

INNOVATION CORRELATED
MULTIVARIATE AUTOREGRESSIVE TIME SERIES
WITH PERIODICALLY CORRELATED PARAMETERS

by

Björn Holmquist

1985:4



**UNIVERSITY OF LUND AND
LUND INSTITUTE OF TECHNOLOGY
DEPARTMENT OF MATHEMATICAL STATISTICS**

LUND, SWEDEN

TYP AV DOKUMENT ☒ Doktorsavhandling ☐ Examensarbete ☐ Kompendium		☐ Ansökan ☐ Reserapport ☐ Delrapport ☐ Slutrapport	☐ Tidskriftsartikel ☐ Konferensuppsats ☐	DOKUMENTBETECKNING/CODEN LUTFD2(TFMS-3038)/1-107/(1985)
AVDELNING/INSTITUTION Department of Mathematical Statistics Box 118 S-221 00 Lund, Sweden				
FÖRFATTARE Björn Holmquist				
DOKUMENTTITEL OCH UNDERTITEL INNOVATION CORRELATED MULTIVARIATE AUTOREGRESSIVE TIME SERIES WITH PERIODICALLY CORRELATED PARAMETERS				
SAMMANFATTNING We study multivariate autoregressive time series with parameters which vary in time in a periodic but unknown way. Estimates of the unknown parameters are derived using two different approaches. One approach is to regard observations corresponding to parameters at a particular time point within the period. The other approach is to estimate the parameters in the deperiodized process obtained by regarding subsequent periods of the process, and from these derive estimates of the parameters in the original process. The asymptotic distributions for long realizations are given for the derived estimators. Portmanteau statistics are proposed for testing consistency of data with the model. We also study two submodels for which either several processes of the given structure are correlated through the innovation processes or, the subprocesses are correlated only through their regression matrices. Statistics for testing these submodels are given and also the asymptotic distributions of the test statistics.				
NYCKELORD Autoregressive process, periodically varying parameters, portmanteau statistic, simple correlated, matrix factorization				
DOKUMENTTITEL OCH UNDERTITEL – SVENSK ÖVERSÄTTNING AV UTLÄNDSK ORIGINALTITEL				
TILLÄMPNINGSOMRÅDE				
TILLÄMPNINGSNYCKELORD				
UTGIVNINGSDATUM år 1985 mån 04		ANTAL SID 107		SPRÅK ☐ svenska ☒ engelska ☐ annat
ÖVRIGA BIBLIOGRAFISKA UPPGIFTER Supported in part by Swedish Natural Science Research Council under Contract No G-UR 4567-100.				ISSN
				ISBN
				PRIS
I, the undersigned, being the copyright owner of the abstract, hereby grant to all reference source permission to publish and disseminate the abstract. Date Signature				

1985-04-29

Björn Holmquist

CONTENTS

	Page
1. Introduction	1
2. The processes and notations	4
3. Stationarity	9
4. Parameter estimation in the general model	15
4.1. Estimation in PMAR processes	15
4.1.1. Least Gramian estimates	15
4.1.2. Yule Walker estimates	19
4.2. Estimation in PMGAR processes	27
4.2.1. Maximum conditional likelihood estimates	29
4.2.2. Approximate maximum likelihood estimates	31
5. Parameter estimation in some reduced models	38
5.1. Time homogeneity	38
5.1.1. Estimation in the P_{R} MAR process	38
5.1.2. Estimation in the P_{I} MAR process	39
5.1.3. Estimation in the P_{RI} MAR process	41
5.2. Reduced correlation structure	42
5.2.1. Estimation in the PSMAR process	42
5.2.2. Estimation in the PBMAR process	46
5.2.3. Estimation in the PSBMAR process	47
6. Diagnostic checking	48
6.1. Known regression parameters	48
6.2. Unknown regression parameters	54
7. Hypothesis testing	57
7.1. Time homogeneity	58
7.1.1. Homogeneous regression in the PMAR process	58
7.1.2. Homogeneous innovations in the PMAR process	58
7.1.3. Time homogeneity in the PMAR process	59
7.2. Reduced correlation structure	61
7.2.1. Simple correlation in the PMAR process	62
7.2.2. Block diagonal innovation covariances in the PMAR process	62
7.2.3. Uncorrelated subprocesses in the PMAR process	64
8. Prediction	67
8.1. Short time prediction	67
8.2. Long time prediction	67

REFERENCES

APPENDIX A: Proof of Lemma 3.1	73
APPENDIX B: Distributions of some compositions and decompositions	74
APPENDIX C: Proofs of Theorems 4.3, 4.5 and 4.6	92
APPENDIX D: Proof of Theorem 4.7	97
APPENDIX E: Proofs of Theorems 6.1 and 6.2	100
APPENDIX F: Proofs of Theorems 6.3 and 6.4	104
APPENDIX G	106

1. INTRODUCTION

The analysis of real time series data using an autoregressive model is a tractable method from several points of view. Firstly, in many situations the model can be given a clear physical interpretation; secondly, it can be successfully fitted to data in many practical applications; thirdly, the theory for such models is well developed making estimation, hypothesis testing and prediction possible; fourthly, implementation of computer algorithms for autoregressive models is being developed which makes data analysis under these models a matter of routine.

The linear autoregressive model used in an earlier stage has in recent years been followed by more complicated autoregressive models in order to fill gaps where the traditional model does not fit suitably. Bilinear autoregressive time series models were introduced by Granger and Anderson (1978) in order to model situations where the data were grossly non-linear. Other more general non-linear autoregressions have been considered by Jones (1978) who obtained results for a first order non-linear autoregressive process. Threshold autoregressions considered by Tong (1978) and Tong and Lim (1980) take into account the non-linearity in the data by modelling different linear autoregressions to different subsets of the data by a procedure determined by the data themselves.

Linear autoregressive models with time dependent parameters have also been considered in order to take the nonlinearity into account. These models have a more or less physical interpretation. Osaki (1980) considers a process with parameters which depend exponentially on time and decay. Subba Rao (1970) discusses a linear autoregressive process with time dependent parameters but without any structure of this dependence. Anderson (1978) discusses estimation and hypothesis testing in such processes when several realisations of the process are available. These processes with time varying parameters can be seen as an introduction to the linear autoregressive processes with random coefficients. These latter

processes are obtained by letting the dependence of the parameters of time be random instead of deterministic. The structure of this randomness can be of different types. Ledolter (1980) states the structure by letting the parameters follow a random walk over time, while the approach of Nicholls and Quinn (1982), Conlisk (1974), (1976) and Andel (1976) is to let the parameters over time be independent variables which are independent of the innovation process in the autoregressive process. The structure of the time dependence of the parameters is here given in terms of the first two moments of the parameter process.

Jones and Brelsford (1967) consider a scalar linear autoregressive process where the parameters have a deterministic periodic behaviour; see also Pagano (1978) and Troutman (1979). They try to include periodic behaviour, which occurs in many physical data series, as for instance in hydrology and meteorology due to changes over the year. In hydrology a simple version of this model is known as the Thomas-Fiering model (Thomas and Fiering (1962)) and is used for generating riverflow like sequences.

An advantage of using autoregressive schemes with periodically varying parameters over integrated autoregressive schemes (Box and Jenkins (1976)) is that the interpretation is clearer. Autoregressive schemes can easily be accepted in many situations since it is quite easy to argue that the value "today" depends on the values "yesterday" and "before yesterday" and that the deterministic function may be disturbed by an error. For seasonally varying data it may be easy to accept that the dependence on "yesterday" may vary from one time of "year" to another, but from "year" to "year" the dependence is typical for the specific "season". For integrated autoregressive schemes the dependence concerns for example the 12'th or the 365'th difference of the regarded process (for monthly and daily data) and it is not so easily grasped, for example, what effect a change in the parameters in the 12'th difference autoregressive scheme have on the behaviour of the regarded process.

In this paper we investigate the multivariate periodically correlated autoregressive process and extend some results given by Pagano (1978) to this case. We derive estimates of the parameters which determine the process, and give the asymptotic behaviour of these estimates when the length of the realization of process increases.

When considering the interrelation between two or more datasets, the generalization of the multivariate linear autoregressive process to more complicated models is almost identical to that of univariate processes. Nicholls and Quinn (1982) treat the random coefficient linear autoregressive process.

The approach of Risager (1981) for analysing interrelation is to consider several scalar linear autoregressive processes for which the innovations of the processes at each time instant are correlated. The simple correlated autoregressive process, as defined by Risager, is a special case of the multivariate autoregressive process. We consider this special case of the multivariate periodically correlated autoregressive process and some of the results of Risager are extended to the general case.

2. THE PROCESSES AND NOTATIONS

For $y_i \in \mathbb{R}^{m,n}$ we denote by $(y_1: \dots: y_p)$ the m by np matrix with the $((j-1)m+k)$ 'th column equal to the k 'th column of y_j , and for any m by n matrix X with columns $x_1, \dots, x_n$ we denote by $\text{vec}(X)$ the column vector in $\mathbb{R}^{mn}$ defined by

$$\text{vec}(X) = (x_1^T: \dots: x_n^T)^T.$$

where for any matrix A , we denote by A^T the transpose of A .

We will use the notation $A \otimes B$ for the Kronecker product (ab_{ij}) of A and $B = (b_{ij})$, $A^{<2>}$ for $A \otimes A$ and δ_{ij} for Kronecker's delta which is 1 if $i=j$ and zero otherwise.

We shall define the processes treated in this paper.

DEFINITION 2.1. A process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ is a periodic multivariate (Gaussian) autoregression of period d and order $p_d^T = (p_1, \dots, p_d)$ if for all $k \in \mathbb{Z}$,

$$(2.1) \quad y_k = \sum_{i=1}^{p_k} \phi_i(k) y_{k-i} + e_k.$$

Here the sequence of errors $\{e_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ are zero mean (mutually independent normally distributed) random vector variables and $E(e_i e_k^T) = \delta_{ik} \Sigma(k)$, $p_k = p_{k+d}$, $\Sigma(k) = \Sigma(k+d)$ and the m -square matrices $\phi_i(k)$ satisfy $\phi_i(k) = \phi_i(k+d)$, $i=1, \dots, p_k$. □

We denote such a process $\text{PM}(G)\text{AR}(p_d^T)$.

Univariate processes ($m=1$) of the same structure have been treated by Jones and Brelsford (1967), Pagano (1978) and Troutman (1979), among others. In case we assume mutually independent and normally distributed innovations $\{e_k\}$, we denote the process $\text{PGAR}(p_d^T)$.

If $d=1$, we exclude the P in the notation and then have the ordinary (multivariate) autoregressive process. We also call these processes aperiodic PMAR and PMGAR, respectively.

Risager (1981) gives a definition of what he calls a simple correlated autoregressive process.

DEFINITION 2.2. (Risager (1982, p. 138)) A simple correlated process of degree $\underline{q}_m = (q_1, \dots, q_m)^T$ is a Gaussian stochastic process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ with the following properties:

$$\begin{pmatrix} y_{1,k} - \phi_{1,1}y_{1,k-1} - \dots - \phi_{q_1,1}y_{1,k-q_1} \\ \vdots \\ y_{m,k} - \phi_{1,m}y_{m,k-1} - \dots - \phi_{q_m,m}y_{m,k-q_m} \end{pmatrix} = e_k.$$

Here $y_{i,k}$ is the i 'th element of y_k and $\{e_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ is a sequence of mutually independent Gaussian vector variables with mean 0 and $E(e_k e_k^T) = \Sigma$. □

We shall call such a process a simple correlated Gaussian autoregression and denote it $\text{SGAR}(\underline{q}_m)$. If the G in the notation is excluded, we relax the assumption of normality and just assume that the innovations are uncorrelated.

It is easily seen that a $\text{SAR}(\underline{q}_m)$ -process is a special case of a $\text{MAR}(p)$ process, namely when $\Phi_i(k) = \Phi_i(1) = \text{diag}(\phi_{i,1}, \dots, \phi_{i,m})$. The degree q_j is obtained as $q_j = \max_{1 \leq i \leq p} (i; (\Phi_i(1))_{jj} \neq 0)$, where $(\Phi_i(1))_{jj}$ denotes the (j,j) 'th element of $\Phi_i(1)$.

The periodic version of SAR, denoted $\text{PSAR}(Q)$ where Q is an m by d matrix, is defined by letting $\phi_{i,j}$ vary with time periodically in such a way that $\phi_{i,j}(k) = \phi_{i,j}(k+d)$ and by letting $E(e_k e_k^T) = E(e_{k+d} e_{k+d}^T) = \Sigma(k)$. The PSAR process is a special case of the PMAR process obtained by letting $\Phi_i(k) = \text{diag}(\phi_{i,1}(k), \dots, \phi_{i,m}(k))$.

We now define the multivariate version of simple correlated autoregressive processes.

DEFINITION 2.3. A process $\{y_k = \text{vec}(y_{1,k} : \dots : y_{s,k}) \in \mathbb{R}^m, k \in \mathbb{Z}\}$, where $m = m_1 + \dots + m_s$ and $y_{i,k} \in \mathbb{R}^{m_i}$, is said to be a multivariate simple

correlated autoregression of degree $\underline{q}_s^T = (q_1, \dots, q_s)$ if for all $k \in \mathbb{Z}$,

$$\begin{pmatrix} y_{1,k} - \phi_{1,11} y_{1,k-1} - \dots - \phi_{q_1,11} y_{1,k-q_1} \\ \vdots \\ y_{s,k} - \phi_{1,ss} y_{s,k-1} - \dots - \phi_{q_s,ss} y_{s,k-q_s} \end{pmatrix} = e_k$$

where $\{e_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ are zero mean random vector variables and

$E(e_i e_k^T) = \delta_{ik} \Sigma$ and where $\phi_{i,j}$ is a m_j -square matrix. The subprocesses $\{y_{i,k} \in \mathbb{R}^{m_i}, k \in \mathbb{Z}\}, i=1, \dots, s$ are said to be innovation correlated.

We denote such a process $\text{MSAR}(\underline{q}_s)$. If the errors $\{e_k\}$ are mutually independent normally distributed we call the process by $\text{MSGAR}(\underline{q}_s)$.

Likewise the univariate case, the $\text{MS(G)AR}(\underline{q}_s)$ process is a special case of the $\text{M(G)AR}(p)$ process obtained by letting $\phi_i(k) = \phi_i(1) = \phi_{i,11} \oplus \dots \oplus \phi_{i,ss} \equiv \text{blockdiag}(\phi_{i,11}, \dots, \phi_{i,ss})$ and where q_j is given by the maximum value of i ($1 \leq i \leq p$) such that the j 'th diagonal block of $\phi_i(1)$ is nonzero.

By letting $\phi_{i,jj}$ and Σ in Definition 2.3 vary with time in a periodic way we obtain the periodic version of this process.

DEFINITION 2.4. A process $\{y_k = \text{vec}(y_{1,k} : \dots : y_{s,k}) \in \mathbb{R}^m, k \in \mathbb{Z}\}$ where $m = m_1 + \dots + m_s$ and $y_{i,k} \in \mathbb{R}^{m_i}$, is a periodic simple correlated multi-variate (Gaussian) autoregression of period d and degree $Q = (q_{i,j})$, $i=1, \dots, s, j=1, \dots, d$, if for all $k \in \mathbb{Z}$,

$$(2.2) \quad \begin{pmatrix} y_{1,k} - \phi_{1,11}(k) y_{1,k-1} - \dots - \phi_{q_{1,k},11}(k) y_{1,k-q_{1,k}} \\ \vdots \\ y_{s,k} - \phi_{1,ss}(k) y_{s,k-1} - \dots - \phi_{q_{s,k},ss}(k) y_{s,k-q_{s,k}} \end{pmatrix} = e_k$$

where $\{e_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ are zero mean (mutually independent Gaussian distributed) random vector variables and $E(e_i e_k^T) = \delta_{ik} \Sigma(k)$, $q_{i,k} = q_{i,k+d}$, $\Sigma(k) = \Sigma(k+d)$ and where the m_j -square matrix $\phi_{i,jj}(k)$ satisfies $\phi_{i,jj}(k) = \phi_{i,jj}(k+d)$.

We denote such a process $\text{PMSAR}(Q)$ ($\text{PMSGAR}(Q)$). This process is a

special case of $\text{PM}(\underline{G})\text{AR}(\underline{p}_d^T)$ obtained by letting $\phi_i(k) = \bigoplus_{j=1}^s \phi_{i,jj}(k)$
 $\equiv \phi_{i,11}(k) \oplus \dots \oplus \phi_{i,ss}(k)$.

The notion of simple correlation refers to the process $\{y_k, k \in \mathbb{Z}\}$ as a whole; its intercorrelation structure is meant to be simple when the regression matrices $\phi_i(k), k=1, \dots, d, i=1, \dots, p_k$ are block diagonal. The subprocesses $y_{i,k}, i=1, \dots, s$ of y_k are said to be innovation correlated since they are correlated (only) through the innovation process.

In a situation where $\{y_{1,k}\}$ is the process of interest, one might ask what is the gain obtained by modelling $\{y_{1,k}\}$ together with $\{y_{2,k}, \dots, y_{s,k}\}$ in a PMSAR process rather than modelling $\{y_{1,k}\}$ solely in a PMAR process. Since the regression matrices are block diagonal, the values of $y_{2,k-j}, \dots, y_{s,k-j}, j=1, 2, \dots$ do not affect $y_{1,k}$. Further, since the subprocess $\{y_{1,k}\}$ in a PMSAR process is a PMAR process it seems non-informative to model the PMSAR process $\{\text{vec}(y_{1,k} : \dots : y_{s,k}), k \in \mathbb{Z}\}$ instead of just modelling the PMAR process $\{y_{1,k}, k \in \mathbb{Z}\}$. In prediction situations, accuracy can sometimes be gained by modelling $\{y_{1,k}, k \in \mathbb{Z}\}$ in an PMSAR process. If the parameters are known and moreover we assume that at a given time point k , $y_{i,k-j}, i=1, \dots, s, j=1, 2, \dots$ and $y_{2,k}, \dots, y_{s,k}$ are known and we want to predict the outcome of $y_{1,k}$, then $e_{2,k}, \dots, e_{s,k}$ can be calculated. Since $e_{1,k}$ is correlated with these in a PMSAR process, a better prediction may hopefully be obtained by predicting conditionally on the outcome of $e_{2,k}, \dots, e_{s,k}$. A similar argument holds when the parameters are unknown and we predict conditionally on the estimated errors $\hat{e}_{2,k}, \dots, \hat{e}_{s,k}$.

A contradiction in some sense to the simple correlated process arises when the innovations of the subprocesses are uncorrelated. This is the case when the covariance matrices $\Sigma(k)$ are block diagonal with the i 'th diagonal block of size m_i -square. We denote such a process by $\text{PMBAR}(\underline{p}_j^T)$ where the sizes of the blocks should be understood from the context.

Following Gladyshev (1961) we define a multivariate periodically correlated time series as follows.

DEFINITION 2.5. A process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ is said to be periodically correlated of period d if $E(y_i y_j^T) = E(y_{i+d} y_{j+d}^T)$. □

We shall use the notation $\Gamma_{ij}(k)$ for $E(y_{k-i} y_{k-j}^T)$ and in this notation $\{y_k\}$ is periodically correlated if

$$(2.3) \quad \Gamma_{ij}(k) = \Gamma_{ij}(k+d).$$

3. STATIONARITY

For the analysis of stationarity of PMAR processes, the following definition is useful.

DEFINITION 3.1. For a process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$, the process $\{x_k \in \mathbb{R}^{md}, k \in \mathbb{Z}\}$ defined by

$$(3.1) \quad x_k = \text{vec}(y_{(k-1)d+1} : \dots : y_{kd})$$

is called the corresponding deperiodized process. □

The relation between periodic autocorrelated sequences $\{y_k\}$ and sequences $\{x_k\}$ given by (3.1) is given in the next theorem.

THEOREM 3.1. The process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ is a $\text{PMAR}(p_d^T)$ process if and only if the deperiodized process $\{x_k \in \mathbb{R}^{md}, k \in \mathbb{Z}\}$ given by (3.1) is a $\text{MAR}(p)$ process with $p = \max_{1 \leq i \leq d} [(p_i - i)/d] + 1$. □

The theorem is an immediate generalization to the multivariate case of a result of Pagano (1978, Theorem 1) and is proven in the same spirit. We shall need the following lemma, which is a generalization of the modified Cholesky decomposition of a non-negative definite matrix (positive definite (p.d.) or positive semi-definite (p.s.d.)). The proof of the lemma is deferred to Appendix A.

Let $\underline{m}_d = (m_1, \dots, m_d)$ and denote by $L_{\underline{m}_d}^{+,+} (L_{\underline{m}_d}^+)$ the space of $m_1 + \dots + m_d$ -square blocked matrices with the (i,j) 'th block of size m_i by m_j being zero if $i > j$ ($i < j$) and with diagonal blocks L_{ii} , $i=1, \dots, d$, such that $L_{ii}^T L_{ii}$ is p.d. Matrices of this structure are referred to as block upper (lower) triangular matrices. Denote by $L_{\underline{m}_d}^{+,I} (L_{\underline{m}_d}^{I,+})$ the subspace of $L_{\underline{m}_d}^{+,+} (L_{\underline{m}_d}^+)$ with identity matrices as diagonal blocks. We call matrices of this structure block unit upper (lower) triangular matrices.

LEMMA 3.1. (Block modified Cholesky decomposition) Let $W = (W_{ij})$ be a non-negative definite $(m_1 + \dots + m_d)$ -square matrix with block W_{ij} of order

m_i by m_j . Then there exists a decomposition $W = LDL^T$ where $L = (L_{ij}) \in L_{m_d}^I$ and where $D = \bigoplus_{i=1}^d D_{ii}$ is a block diagonal matrix with D_{ii} being m_i -square non-negative definite. The blocks of L and D are given by

$$(3.2) \quad D_{jj} = W_{jj} - \sum_{k=1}^{j-1} L_{jk} D_{kk} L_{jk}^T, \quad j=1, \dots, d$$

$$(3.3) \quad L_{ij} = (W_{ij} - \sum_{k=1}^{j-1} L_{ik} D_{kk} L_{jk}^T) D_{jj}^{-1}, \quad i > j, \quad i, j=1, \dots, d$$

for some g-inverse D_{jj}^{-} of D_{jj} . If W is positive definite the decomposition is unique, in which case D_{jj} is positive definite and given by (3.2) and L_{ij} is given by (3.3) with the inverse of D_{jj} replacing D_{jj}^{-} .

PROOF OF THEOREM 3.1. Assume that $\{y_k \in R^m, k \in Z\}$ is a PMAR(p_d^T) process. Then we can write the process $\{x_k \in R^{md}, k \in Z\}$,

$$(3.4) \quad \Lambda^{-1} x_k = \sum_{i=1}^p F_i x_{k-i} + v_k$$

where

$$(3.5) \quad \Lambda^{-1} = \begin{pmatrix} I_m & 0 & 0 & \dots & 0 & 0 \\ -\Phi_1(2) & I_m & 0 & \dots & 0 & 0 \\ -\Phi_2(3) & -\Phi_1(3) & I_m & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ -\Phi_{d-1}(d) & -\Phi_{d-2}(d) & -\Phi_{d-3}(d) & \dots & -\Phi_1(d) & I_m \end{pmatrix}$$

and

$$(3.6) \quad F_i = \begin{pmatrix} \Phi_{id}(1) & \Phi_{id-1}(1) & \dots & \Phi_{(i-1)d+1}(1) \\ \Phi_{id+1}(2) & \Phi_{id}(2) & \dots & \Phi_{(i-1)d+2}(2) \\ \vdots & \vdots & & \vdots \\ \Phi_{(i+1)d-1}(d) & \Phi_{(i+1)d-2}(d) & \dots & \Phi_{id}(d) \end{pmatrix}$$

where $\Phi_i(k) = 0$ if $i > p_k$ and $v_k = \text{vec}(e_{1+(k-1)d} \dots e_{kd})$. We have $E(v_k) = 0$ and $E(v_i v_k^T) = \delta_{ik} \bigoplus_{j=1}^d \Sigma(j)$. Clearly $\Sigma(1) \oplus \dots \oplus \Sigma(d)$ is p.d. since so are all $\Sigma(i)$, $i=1, \dots, d$. Thus

$$(3.7) \quad x_k = \sum_{i=1}^p \Psi_i x_{k-i} + u_k$$

where $\Psi_i = \Lambda F_i$, $i=1, \dots, p$ and $u_k = \Lambda v_k$. Here the quantities $u_k \in \mathbb{R}^{md}$, $k \in \mathbb{Z}$, are uncorrelated with zero mean and $E(u_k u_k^T) = \Lambda \left(\bigoplus_{j=1}^d \Sigma(j) \right) \Lambda^T \equiv \Theta$ is positive definite.

On the other hand, assume that the process $\{x_k \in \mathbb{R}^{md}, k \in \mathbb{Z}\}$ satisfies (3.7) for some matrices Ψ_i , $i=1, \dots, p$, where $\{u_k\}$ is an uncorrelated sequence with p.d. covariance matrix Θ . Then, by Lemma 3.1, there exists a unique decomposition $\Theta = \Lambda \Delta \Lambda^T$ where Λ is block unit lower triangular and Δ is a p.d. block diagonal matrix. Thus (3.4) is satisfied with $F_i = \Lambda^{-1} \Psi_i$ and where the errors $v_k = \Lambda^{-1} u_k$, $k \in \mathbb{Z}$, are uncorrelated with positive definite block diagonal covariance matrix Δ . Hence the process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ satisfies the relation in Definition 2.1 for some matrices $\Phi_i(k)$ and $\Sigma(k)$ which can be obtained from the blocks of Λ^{-1} and F_i given by (3.5) and (3.6) and from the diagonal blocks of Δ . $\square$

Remark 3.1. If the process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ is a PMAR process of period d it has also period nd for positive integers n , and thus we can always find a process x_k given in Theorem 3.1 such that p in that theorem is equal to 1. $\square$

The exact relation between periodically correlated sequences $\{y_k\}$ and second order stationarity of the corresponding deperiodized process is given by Gladyšev (1961). The following version will be needed here.

THEOREM 3.2. (Gladyšev (1961)) The process $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ is periodically correlated if and only if the corresponding deperiodized process $\{x_k \in \mathbb{R}^{md}, k \in \mathbb{Z}\}$ given by (3.1) is second order stationary. $\square$

Remark 3.2. We note that by the result of Gladyšev (Theorem 3.2), a PMAR process is periodically correlated if and only if its deperiodized counterpart given in Theorem 3.1 is a second order stationary MAR process. If the deperiodized process is second order stationary, this implies periodic behaviour of $\Gamma_{ij}(k)$ in k as given by (2.3), in addition to that of

$\phi_i(k)$, $i=1, \dots, p_k$ and $\Sigma(k)$. □

Stationarity of the process given by (3.7) requires that the roots of the equation

$$\left| \sum_{i=0}^p \lambda^{p-i} \psi_i \right| = 0$$

($\psi_0 = -I_m$) are less than 1 in absolute value (see e.g. Anděl (1971)). By premultiplication by $|\Lambda^{-1}|$ we obtain the criterion for stationarity of the process $\{y_k\}$ given by (2.1): if the roots of the equation

$$(3.8) \quad \left| \sum_{i=0}^p \lambda^{p-i} F_i \right| = 0$$

($F_0 = -\Lambda^{-1}$) are less than 1 in absolute value, then the PMAR process is stationary.

Utilizing the fact noted in Remark 3.1, we let $d' = n'd$ for some integer n' such that $d' \geq \max_{1 \leq k \leq d} p_k$ and thus $\phi_i(k) = \phi_i(k+nd')$ and $\Sigma(k+nd') = \Sigma(k)$ for all integers n . This value d' is thus a candidate for being the period of the PMAR process although it is not necessarily the least integer with the above given properties.

We write equation (2.1) in the following companion form:

$$\text{vec}(y_{k-d'+1} : \dots : y_k) = \underline{\Phi}(k) \text{vec}(y_{k-d'} : \dots : y_{k-1}) + \text{vec}(0 : \dots : 0 : e_k)$$

where

$$\underline{\Phi}(k) = \begin{pmatrix} 0 & | & I_{m(d'-1)} \\ \hline \phi_{d'}(k) : \dots : \phi_1(k) \end{pmatrix}$$

and $\phi_i(k) = 0$ for $i > p_k$. By using this equation recursively we obtain

$$\begin{aligned} \text{vec}(y_{k-d'+1} : \dots : y_k) &= A_{d'}(k) \text{vec}(y_{k-2d'+1} : \dots : y_{k-d'}) \\ &\quad + \sum_{j=0}^{d'-1} A_j(k) \text{vec}(0 : \dots : 0 : e_{k-j}) \end{aligned}$$

where $A_j(k) = \underline{\Phi}(k) \underline{\Phi}(k-1) \dots \underline{\Phi}(k-j+1)$ and $A_0(k) = I_{md'}$.

Let $z_k = \text{vec}(y_{(k-1)d'+1} : \dots : y_{kd'})$, then since $A_j(kd')$ does not

depend on k for $j=1, \dots, d'$, we have

$$z_k = Az_{k-1} + w_k$$

where $A = A_{d', (d')}$ and $w_k = \sum_{j=0}^{d'-1} A_j(kd') \text{vec}(0: \dots : 0: e_{kd'-j})$, $k \in \mathbb{Z}$, is an uncorrelated sequence with mean zero and dispersion

$$\underline{\Theta} = \sum_{j=0}^{d'-1} A_j(d') (0_{(d'-1)m, (d'-1)m} \oplus \Sigma(d'-j)) A_j(d')^T.$$

We observe that if $\max_{1 \leq k \leq d} p_k \leq d$, we can choose $d' = d$ and in this case $z_k = x_k$ where x_k is given in Definition 3.1. The value of p , given in Theorem 3.1, is then equal to 1, and $A = \Psi_1$, $w_k = u_k$ and $\underline{\Theta} = \Theta$.

If $\max_{1 \leq k \leq d} p_k > d$, then d' has to be chosen larger than d , $d' = n'd$ say, for some positive integer n' . In this case

$$(3.9) \quad z_k = \text{vec}(x_{n', (k-1)+1} : \dots : x_{n', k}),$$

and

$$A = \begin{pmatrix} 0 & I_{md(n'-1)} \\ \Psi_{n'} : \Psi_{n', -1} : \dots : \Psi_1 \end{pmatrix}$$

where $\Psi_k = 0$ if $k > p$, and

$$\underline{\Theta} = \begin{pmatrix} 0 & 0 \\ 0 & \Theta \end{pmatrix}.$$

Obviously we can take $n'=p$ as given in Theorem 3.1.

We have

$$E(z_k z_k^T) = A E(z_{k-1} z_{k-1}^T) A^T + \underline{\Theta},$$

and with $W_{n'} = E(z_k z_k^T)$ which is independent of k , we have

$$W_{n'} = A W_{n'} A^T + \underline{\Theta}.$$

This equation has the solution

$$(3.10) \quad W_{n'} = \sum_{j=0}^{\infty} A^j \underline{\Theta} A^{Tj}$$

for the dispersion matrix $W_{n'}$. Further

$$E(z_{k+s} z_k^T) = A^s E(z_k z_k^T) = A^s W_{n'} = \sum_{j=0}^{\infty} A^{j+s} \underline{\Theta} A^{Tj}.$$

We shall need the following result which for $md=1$ is identical to Lemma 5.5.5 in Anderson (1971). The proof of the lemma is immediately generalized to cover the present case.

LEMMA 3.2. If θ is p.d., then W_n , given by (3.10) is p.d.

4. PARAMETER ESTIMATION IN THE GENERAL MODEL

In this section we shall derive four different estimates of the parameters $\phi_i(k)$ and $\Sigma(k)$, $k=1, \dots, d$, $i=1, \dots, p_k$. Two of the estimates, the least Gramian estimates and the Yule-Walker estimates apply to the general PMAR-process, while the other two, the maximum conditional likelihood estimates and the approximate maximum likelihood estimates, will be investigated for the PMGAR-process.

We assume that the quantities d and $p_1, \dots, p_d$ are known and that we are given a sample of length $n+q+1$ labelled $y_{-q}, \dots, y_0, y_1, \dots, y_n$ where $q = \max_{1 \leq j \leq d} (p_j - j)$.

We shall use the following distributional notations.

Vector normal distribution: $x = (x_i)$ (in R^m) $\in N_m(\mu, \Sigma)$ if the ch.f. at $t \in R^m$ is given by $\exp(it^T \mu - \frac{1}{2} t^T \Sigma t)$ and the density for $\Sigma > 0$ is given by

$$(2\pi)^{-\frac{1}{2}m} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu))$$

with respect to $(dx) \equiv \prod_i dx_i$.

Matric normal distribution: $X = (x_{ij})$ (in $R^{m,n}$) $\in N_{m,n}(M, \Sigma, \Gamma)$ if the ch.f. at $T \in R^{m,n}$ is given by $\exp(\text{tr}(iT^T M - \frac{1}{2} \Sigma T \Gamma T^T))$ and the density for $\Sigma > 0$ and $\Gamma > 0$ is given by

$$(2\pi)^{-\frac{1}{2}mn} |\Sigma|^{-\frac{1}{2}n} |\Gamma|^{-\frac{1}{2}m} \exp(-\frac{1}{2} \text{tr}(\Sigma^{-1} (X-M) \Gamma^{-1} (X-M)^T))$$

with respect to $(dX) \equiv \prod_i \prod_j dx_{ij}$.

4.1. ESTIMATION IN PMAR PROCESSES.

4.1.1. Least Gramian estimates

Firstly assume that $\phi(1), \dots, \phi(d)$ are known but $\Sigma(1), \dots, \Sigma(d)$ are unknown. Let $[x]$ denote the integer part of x and define $N = [n/d]$ and $N_k = N + 1_{\{k \leq n - Nd\}}$. Define

$$Y_k = (y_k : y_{k+d} : \dots : y_{k+(N_k-1)d})$$

$$y'_k = \text{vec}(y_{k-1} : \dots : y_{k-p_k}),$$

$$Y'_k = (y'_k : y'_{k+d} : \dots : y'_{k+(N_k-1)d})$$

and

$$\Phi(k) = (\Phi_1(k) : \dots : \Phi_{p_k}(k)).$$

Since the errors $e_k = y_k - \Phi(k)y'_k$, have the dispersion matrix $\Sigma(k)$ for $k = 1, \dots, n$, unbiased estimators of $\Sigma(k)$, $k = 1, \dots, d$, are given by

$$(4.1) \quad (Y_k - \Phi(k)Y'_k)(Y_k - \Phi(k)Y'_k)^T / N_k \equiv Q_{kk}(\Phi(k)) / N_k$$

where $Q_{ij}(A)$ is defined by

$$Q_{ij}(A) = (Y_i - AY'_i)(Y_j - AY'_j)^T.$$

Next assume that also $\Phi(1), \dots, \Phi(d)$ are unknown. It is natural to use the Gramian matrix $Q_{kk}(\Phi(k))$ to obtain an estimate of $\Phi(k)$. In analogy with least square estimation, we take as an estimate of $\Phi(k)$ the value which "minimizes" the Gramian matrix and this estimate will be called a least Gramian (LG) estimate of $\Phi(k)$. The sense in which the Gramian matrix is minimized is that $Q_{kk}(A)$ is not greater than $Q_{kk}(B)$ if $Q_{kk}(B) - Q_{kk}(A)$ is non-negative definite. In this case we write $Q_{kk}(B) \geq_m Q_{kk}(A)$ or $Q_{kk}(A) \leq_m Q_{kk}(B)$ when $Q_{kk}(B)$ (and $Q_{kk}(A)$) is m-square. Hence $\hat{A}$ minimizes $Q_{kk}(A)$ if $Q_{kk}(\hat{A}) \leq_m Q_{kk}(\hat{A} + \Delta)$ for any matrix Δ of the same size as $\hat{A}$. This is equivalent to that $(Y_k - \hat{A}Y'_k)Y_k^T \Delta^T = 0$ for all Δ , which is the same as $(Y_k - \hat{A}Y'_k)Y_k^T = 0$. Hence $Q_{kk}(\hat{A} + \Delta) = Q_{kk}(\hat{A}) + \Delta Y'_k Y_k^T \Delta^T$ and since $\Delta Y'_k Y_k^T \Delta^T$ is non-negative definite, the j 'th eigenvalue of $Q_{kk}(\hat{A} + \Delta)$ is at least as large as the j 'th eigenvalue of $Q_{kk}(\hat{A})$ and is non-negative. This implies that

$$\text{tr}_j(Q_{kk}(\hat{A} + \Delta)) \geq \text{tr}_j(Q_{kk}(\hat{A}))$$

for $j=1, \dots, m$, where $\text{tr}_j(C)$ is the j 'th order elementary symmetric function of the eigenvalues of the m-square matrix C . Especially

$$\text{tr}(Q_{kk}(\hat{A} + \Delta)) \geq \text{tr}(Q_{kk}(\hat{A}))$$

and thus $\hat{A}$ is also a least (trace) norm (LN) estimate. Also

$$|Q_{kk}(\hat{A} + \Delta)| \geq |Q_{kk}(\hat{A})|,$$

where $|C|$ denotes the determinant of C , and $\hat{A}$ is here referred to as a least generalized (sum of) squares (LGS) estimate. The notion of generalized sum of squares is in analogy with the notion of generalized variance, Wilks (1932), being the determinant of a dispersion matrix.

DEFINITION 4.1. Let $F(k)$ be a parameter space for $\phi(k)$. A set $\{\hat{\phi}^{LG}(k) \in F(k) \text{ a.s., } k=1, \dots, d\}$ of random variables such that

$$\begin{aligned} (Y_1 - \hat{\phi}^{LG}(1)Y'_1)(Y_1 - \hat{\phi}^{LG}(1)Y'_1)^T &\leq_m (Y_1 - \phi(1)Y'_1)(Y_1 - \phi(1)Y'_1)^T \\ &\vdots \\ (Y_d - \hat{\phi}^{LG}(d)Y'_d)(Y_d - \hat{\phi}^{LG}(d)Y'_d)^T &\leq_m (Y_d - \phi(d)Y'_d)(Y_d - \phi(d)Y'_d)^T \end{aligned}$$

all holds a.s. for all $\phi(k) \in F(k)$, is called a set of least Gramian estimators of $\{\phi(k) \in F(k), k=1, \dots, d\}$ in the PMAR model. A class of least Gramian estimator $\hat{\Sigma}^{LG}(k)$ of $\Sigma(k)$ is defined by

$$(4.2) \quad \hat{\Sigma}^{LG}(k) = (Y_k - \hat{\phi}^{LG}(k)Y'_k)(Y_k - \hat{\phi}^{LG}(k)Y'_k)^T / f_k$$

where f_k is a suitably chosen number.

THEOREM 4.1. Let $F(k)$ be the range space of the PMAR process given by

(2.1), then there exists a set $\{\hat{\phi}^{LG}(k) \in F(k) \text{ a.s. } k=1, \dots, d\}$ of LG estimators for the PMAR process. The estimators can be taken to be

$$(4.3) \quad \hat{\phi}^{LG}(k) = Y_k Y'_k{}^T (Y'_k Y'_k{}^T)^-, \quad k=1, \dots, d$$

where $(Y'_k Y'_k{}^T)^-$ is an arbitrary g-inverse of $Y'_k Y'_k{}^T$.

PROOF. Write $\phi(k) = \hat{\phi}^{LG}(k) - \Delta$, where $\hat{\phi}^{LG}(k)$ is given by (4.3). Then since $(Y_k - \hat{\phi}^{LG}(k)Y'_k)(\Delta Y'_k)^T = 0$, it follows that

$$\begin{aligned} (Y_k - \phi(k)Y'_k)(Y_k - \phi(k)Y'_k)^T &= (Y_k - \hat{\phi}^{LG}(k)Y'_k)(Y_k - \hat{\phi}^{LG}(k)Y'_k)^T \\ &\quad + \Delta Y'_k Y'_k{}^T \Delta^T \end{aligned}$$

and thus, since $\Delta Y'_k Y'_k{}^T \Delta^T \geq_m 0$, we have that

$$\begin{aligned} (Y_k - \phi(k)Y'_k)(Y_k - \phi(k)Y'_k)^T &\geq (Y_k - \hat{\phi}^{LG}(k)Y'_k)(Y_k - \hat{\phi}^{LG}(k)Y'_k)^T \\ &= Y_k(I_{N_k} - Y_k'^T(Y_k'Y_k')^{-1}Y_k')Y_k^T \end{aligned}$$

which proves the theorem. □

We note that

$$\hat{\phi}^{LG}(k)Y'_k = Y_k Y_k'^T (Y_k' Y_k')^{-1} Y'_k$$

is invariant for any version of $(Y_k' Y_k')^{-1}$, so that for smoothing purposes using $\hat{\phi}^{LG}(k)Y'_k$ as a smoothed version of Y_k , the particular choice of $\hat{\phi}^{LG}(k)$ is immaterial. Also the choice of g-inverse is immaterial for the estimator of $\Sigma(k)$ given by (4.2).

The choice of f_k in (4.2) is somewhat arbitrary. A particular choice can be motivated by comparison with multivariate Gauss-Markov regression theory which we refer to as the multivariate linear normal model (MLNM). In this model we let $y_k = B(k)z_k + e_k$ where $\{e_k\}_1^\infty$ is a sequence, independently distributed with $e_k \in N_m(0, \Sigma(k))$, z_k is a mp_k -dimensional fixed and known vector, $p_{k+d} = p_k$ and $B(k)$ is an m by mp_k unknown matrix, such that $B(k+d) = B(k)$. Hence

$$Y_k \in N_{m, N_k}(B(k)Z_k, \Sigma(k), I_{N_k})$$

where $Y_k = (y_k : y_{k+d} : \dots : y_{k+(N_k-1)d})$ and $Z_k = (z_k : z_{k+d} : \dots : z_{k+(N_k-1)d})$.

The LG estimate of $B(k)$ based on Y_k and Z_k is

$$B(k)^* = Y_k Z_k^T (Z_k Z_k^T)^{-1}$$

and an estimate of $\Sigma(k)$ is the normalized residual Gramian $\Sigma(k)^*$ given by

$$\begin{aligned} \Sigma(k)^* &= (Y_k - B(k)^* Z_k)(Y_k - B(k)^* Z_k)^T / f_k \\ &= Y_k (I_{N_k} - Z_k^T (Z_k Z_k^T)^{-1} Z_k) Y_k^T / f_k = (\Sigma_{ij}(k)^*) \quad i, j = 1, \dots, s \end{aligned}$$

where f_k is a suitably chosen number. Here $\Sigma(k)^*$ is distributed as $W_m(N_k - r_k, \Sigma(k)/f_k)$ where $f_k = \text{rank}(I_{N_k} - Z_k^T (Z_k Z_k^T)^{-1} Z_k) = N_k - r_k$ and

$r_k = \text{rank}(Z_k) = \text{rank}(Z_k^T(Z_k Z_k^T)^{-1}Z_k) = \text{tr}(Z_k^T(Z_k Z_k^T)^{-1}Z_k)$. A criterion for choosing f_k may be unbiasedness of $\Sigma(k)^*$ which gives $f_k = N_k - r_k$.

In the present case the design matrix corresponding to Z_k is Y_k' which is random. However the design Z_k in the MLNM case may be random if it is distributed independently of Y_k . In this case

$$\Sigma(k)^* | r_k \in W_m(N_k - r_k, \Sigma(k)/f_k)$$

and again the choice $f_k = N_k - r_k$ (now being random) yields unbiasedness of $\hat{\Sigma}_k$, and this even unconditionally.

Crucial in the present case is, however, the dependence between Y_k and Y_k' .

In the following we will use the estimate

$$(4.4) \quad \hat{\Sigma}^{LG}(k) = Y_k(I_{N_k} - Y_k'^T(Y_k'Y_k'^T)^{-1}Y_k')Y_k^T/(N_k - r_k^0)$$

where $r_k^0 = \text{rank}Y_k' = \text{tr}(Y_k'^T(Y_k'Y_k'^T)^{-1}Y_k')$.

4.1.2. Yule-Walker estimates

Estimation without deperiodization

In PMAR(p_d^T) processes $\Sigma(k)$ and $\Phi_i(k)$, $i=1, \dots, p_k$, are related via equations which can be regarded as periodic multivariate Yule-Walker equations and are given in the theorem below. These equations are extensions to the multivariate case of results of Pagano (1978, Theorem 2), and to the periodic case of results of Whittle (1963, formulas (11) and (12) when $q=0$) as well as of Hannan (1970, (2.3) p. 327).

THEOREM 4.2. The second order moments $\Gamma_{ij}(k) = E(y_{k-i}y_{k-j}^T)$ of the PMAR(p_d^T) process $\{y_k \in R^m, k \in Z\}$ generated by (2.1) satisfy the following relations.

$$(i) \quad \Gamma_{i0}(k) = \sum_{j=1}^{p_k} \Gamma_{ij}(k) \Phi_j(k)^T, \quad i > 0, k \in Z$$

$$(ii) \quad \Gamma_{00}(k) = \sum_{j=1}^{p_k} \Gamma_{0j}(k) \Phi_j(k)^T + \Sigma(k), \quad k \in \mathbb{Z},$$

$$(iii) \quad \Gamma_{i0}(k) = \sum_{j=1}^{p_{k-i}} \Phi_j(k-i) \Gamma_{i+j,0}(k), \quad i < 0, \quad k \in \mathbb{Z}.$$

PROOF. We have, using (2.1),

$$y_{k-i} y_k^T = y_{k-i} \left(\sum_{j=1}^{p_k} y_{k-j}^T \Phi_j(k)^T + e_k^T \right).$$

Taking expectations on both sides we get

$$\Gamma_{i0}(k) = \sum_{j=1}^{p_k} \Gamma_{ij}(k) \Phi_j(k)^T + E(y_{k-i} e_k^T),$$

where the last term is equal to $\Sigma(k)$ if $i=0$ and zero if $i>0$. We also have, again using (2.1),

$$y_{k-i} y_k^T = \left(\sum_{j=1}^{p_{k-i}} \Phi_j(k-i) y_{k-i-j}^T + e_{k-i}^T \right) y_k^T$$

and thus by taking expectations we obtain, for $i < 0$,

$$\Gamma_{i0}(k) = \sum_{j=1}^{p_{k-i}} \Phi_j(k-i) \Gamma_{i+j,0}(k)$$

$$\text{and for } i=0, \quad \Gamma_{00}(k) = \sum_{j=1}^{p_k} \Phi_j(k) \Gamma_{j0}(k) + \Sigma(k).$$

It will be convenient to introduce the $m p_k$ -square matrix $\Gamma_{**}(k) = E(y_k' y_k'^T)$ which is equal to

$$\Gamma_{**}(k) = \begin{pmatrix} \Gamma_{11}(k) & \dots & \Gamma_{1p_k}(k) \\ \vdots & & \vdots \\ \Gamma_{p_k 1}(k) & \dots & \Gamma_{p_k p_k}(k) \end{pmatrix}.$$

We also introduce

$$\Gamma_{\cdot 0}(k) = E(y_k' y_k'^T) = \Gamma_{0\cdot}(k)^T = \begin{pmatrix} \Gamma_{10}(k) \\ \vdots \\ \Gamma_{p_k 0}(k) \end{pmatrix}$$

which is of order $m p_k$ by m . In these notations the following relations

can be deduced from the results of Theorem 4.2:

$$(4.5) \quad \Gamma_{1,0}(k) = \Gamma_{11}(k)\Phi(k)^T$$

and

$$(4.6) \quad \Gamma_{00}(k) = \Gamma_{01}(k)\Phi(k)^T + \Sigma(k).$$

As in the time independent case it is natural to use these relations for estimating the unknown parameters.

DEFINITION 4.2. Any solutions $\hat{\Sigma}^{YW}(k)$ and $\hat{\Phi}^{YW}(k)$ of the Yule-Walker equations (4.5) and (4.6) with estimates $\hat{\Gamma}_{ij}(k)$ inserted for $\Gamma_{ij}(k)$, $i, j=0, 1, \dots, p_k$, are called Yule-Walker estimates of $\Sigma(k)$ and $\Phi(k)$. $\square$

In general we restrict the class of estimates of $\Gamma_{ij}(k)$ to consistent estimates.

Let $\hat{\Gamma}_{11}(k)$ be the matrix obtained from $\Gamma_{11}(k)$ when $\Gamma_{ij}(k)$ is replaced by $\hat{\Gamma}_{ij}(k)$ and let $\hat{\Gamma}_{1,0}(k)$ and $\hat{\Gamma}_{0,1}(k)$ be defined analogously.

By solving for $\Phi(k)$ and $\Sigma(k)$ in (4.5) and (4.6) we obtain the Yule-Walker estimates $\hat{\Sigma}^{YW}(k)$ and $\hat{\Phi}^{YW}(k)$ based on $\hat{\Gamma}_{ij}(k)$, $i, j \in \{0, \dots, p_k\}$ as

$$\hat{\Phi}^{YW}(k) = \hat{\Gamma}_{0,1}(k)\hat{\Gamma}_{11}(k)^{-}$$

and

$$\begin{aligned} \hat{\Sigma}^{YW}(k) &= \hat{\Gamma}_{00}(k) - \hat{\Gamma}_{0,1}(k)\hat{\Phi}^{YW}(k)^T = \hat{\Gamma}_{00}(k) - \hat{\Phi}^{YW}(k)\hat{\Gamma}_{11}(k)\hat{\Phi}^{YW}(k)^T \\ &= \hat{\Gamma}_{00}(k) - \hat{\Gamma}_{0,1}(k)\hat{\Gamma}_{11}(k)^{-}\hat{\Gamma}_{1,0}(k) \end{aligned}$$

where $\hat{\Gamma}_{11}(k)^{-}$ is an g-inverse of $\hat{\Gamma}_{11}(k)$.

We partition Y'_k into

$$(4.7) \quad Y'_k = \begin{pmatrix} Y_{1,k} \\ \vdots \\ Y_{p_k,k} \end{pmatrix}$$

where the m by N_k matrices $Y_{i,k} = (y_{k-i} : y_{k-i+d} : \dots : y_{k-i+(N_k-1)d})$ for $i=1, \dots, p_k$, can differ slightly from Y_{k-i} depending on the values of N_k

and N_{k-1} . We also introduce $Y_{0,k} \equiv Y_k$.

Consider the following statistics:

$$G_{,0}(k) = Y_k' Y_k^T / N_k = G_{0,}(k)^T,$$

$$G_{ij}(k) = Y_{i,k} Y_{j,k}^T / N_k = G_{ji}(k)^T,$$

for $i, j=0, 1, \dots, p_k$, and

$$G_{,,}(k) = Y_k' Y_k^T / N_k = \begin{pmatrix} G_{11}(k) & \dots & G_{1p_k}(k) \\ \vdots & & \vdots \\ G_{p_k 1}(k) & \dots & G_{p_k p_k}(k) \end{pmatrix}$$

with expectations

$$E(G_{,0}(k)) = \sum_{u=0}^{N_k-1} \Gamma_{,0}(k+ud) / N_k,$$

$$E(G_{ij}(k)) = \sum_{u=0}^{N_k-1} \Gamma_{ij}(k+ud) / N_k$$

and

$$E(G_{,,}(k)) = \sum_{u=0}^{N_k-1} \Gamma_{,,}(k+ud) / N_k$$

respectively. From the relation (2.3) for a periodically correlated PMAR process these expectations reduces to $\Gamma_{,0}(k)$, $\Gamma_{ij}(k)$ and $\Gamma_{,,}(k)$ respectively.

We shall study the Yule-Walker estimates obtained by estimating $\hat{\Gamma}_{ij}(k)$ with $G_{ij}(k)$.

Let $T_{m,m}$ denote the direct product transposition matrix (Holmquist, 1985a) also known as the commutation matrix, cf. e.g. Magnus and Neudecker (1979), and let $I_{m,m} = I_m \otimes I_m = I_{\frac{m}{2}}$ be the direct product identity matrix, (Holmquist, 1985a). We define

$$S_m = \frac{1}{2}(T_{m,m} + I_{m,m}).$$

THEOREM 4.3. The Yule-Walker estimates of $\Phi(k)$ and $\Sigma(k)$, in a PMAR process generated by (2.1), based on the estimate $G_{ij}(k)$ of $\Gamma_{ij}(k)$, are given by

$$\hat{\Phi}^{YW}(k) = Y_k Y_k' (Y_k' Y_k')^{-1}, \quad k=1, \dots, d$$

and

$$\hat{\Sigma}^{YW}(k) = Y_k (I_{N_k} - Y_k' (Y_k' Y_k')^{-1} Y_k) Y_k' / N_k$$

The estimates are Fisher consistent and if the roots of the equation (3.8) are less than one in absolute value and the innovations are i.i.d. then the estimates are almost surely consistent.

Further

$$\sqrt{N_k} (\hat{\Phi}^{YW}(k) - \Phi(k)) \xrightarrow{L} N_{m, mp_k} (0, \Sigma(k), \Gamma_{11}(k)^{-1}) \quad \text{as } n \rightarrow \infty.$$

and

$$\sqrt{N_k} \text{vec}(\hat{\Sigma}^{YW}(k) - \Sigma(k)) \xrightarrow{L} N_m^2 (0, 2S_m \Sigma(k) \langle 2 \rangle S_m)$$

PROOF. See Appendix C. □

Estimation in the deperiodized MAR process

For the corresponding deperiodized MAR process we have the following result.

THEOREM 4.4. The second order moments $T_{ij} = E(x_{k-i} x_{k-j}^T)$ of the deperiodized process $\{x_k \in R^{md}, k \in Z\}$ given by (3.1) satisfy the following relations

$$\begin{aligned} \text{(i)} \quad T_{i0} &= \sum_{j=1}^P T_{ij} \Psi_j^T, \quad i > 0 \\ \text{(ii)} \quad T_{00} &= \sum_{j=1}^P T_{0j} \Psi_j^T + \Lambda \left(\bigoplus_{k=1}^d \Sigma(k) \right) \Lambda^T. \end{aligned}$$

PROOF. We have from (3.7)

$$x_{k-i} x_i^T = x_{k-i} \left(\sum_{j=1}^P x_{k-j}^T \Psi_j^T + u_k^T \right).$$

By taking expectations we get

$$T_{i0} = \sum_{j=1}^P T_{ij} \Psi_j^T + E(x_{k-i} u_k^T).$$

Here $E(x_{k-i}^T u_k^T)$ is equal to $\Lambda(\bigoplus_{k=1}^d \Sigma(k))\Lambda^T$ if $i=0$ and equal to zero if $i>0$.

It will be convenient to introduce the mdp -square matrix $T_{,,} = E(x'_k x'_k{}^T)$ where $x'_k = \text{vec}(x_{k-1} : \dots : x_{k-p})$. Thus

$$T_{,,} = \begin{pmatrix} T_{11} & \dots & T_{1p} \\ \vdots & & \vdots \\ T_{p1} & \dots & T_{pp} \end{pmatrix}.$$

We also introduce

$$T_{,0} = E(x'_k x_k^T) = T_{0,}^T = \begin{pmatrix} T_{10} \\ \vdots \\ T_{p0} \end{pmatrix}$$

which is of order mdp by md . In these notations we derive, from Theorem 4.4, the equations

$$(4.8) \quad T_{,0} = T_{,,}\Psi^T$$

and

$$(4.9) \quad T_{00} = T_{0,}\Psi^T + \Theta$$

where $\Theta = \Lambda(\bigoplus_{k=1}^d \Sigma(k))\Lambda^T$ and $\Psi = (\Psi_1 : \dots : \Psi_p)$.

DEFINITION 4.3. Any solutions $\hat{\Theta}^{YW}$ and $\hat{\Psi}^{YW}$ of the Yule-Walker equations (4.8) and (4.9) with estimates $\hat{T}_{ij}$ inserted for T_{ij} , $i, j=0, 1, \dots, p$, are called Yule-Walker estimates of Θ and Ψ in the deperiodized PMAR process. The deperiodized Yule-Walker estimates $\hat{\Phi}_i^{dYW}(k)$ and $\hat{\Sigma}^{dYW}(k)$ of $\Phi_i(k)$ and $\Sigma(k)$, $i=1, \dots, p$, $k=1, \dots, d$, are obtained from the relations

$$\hat{\Psi}^{YW} = \hat{\Lambda}\hat{F}$$

and

$$\hat{\Theta}^{YW} = \hat{\Lambda}(\bigoplus_{k=1}^d \hat{\Sigma}^{dYW}(k))\hat{\Lambda}^T$$

and (3.5) and (3.6) with $\hat{F} = (\hat{F}_1 : \dots : \hat{F}_p)$, $\hat{\Lambda}^{-1}$ and $\hat{\Phi}_i^{dYW}(k)$ substituted

for $F = (F_1 : \dots : F_p)$, Λ^{-1} and $\phi_i(k)$, $i=1, \dots, p_k$ and $k=1, \dots, d$. $\square$

Let $X = (x_1 : \dots : x_N)$ and $X' = (x'_1 : \dots : x'_N)$. We shall study the Yule-Walker estimates in the deperiodized PMAR process and the deperiodized Yule-Walker estimates of the parameters in the PMAR process obtained from $C_{00} = XX^T/N$, $C_{0'} = XX'^T/N = C_{0'}^T$ and $C_{..} = X'X'^T/N$ which are unbiased estimators of T_{00} , $T_{0'} = T_{0'}^T$ and $T_{..}$ respectively.

We write $X^T = (Y_1^T : \dots : Y_d^T)$, where Y_i is essentially the same matrix as given before, although now of dimension m by $N = \min(N_1, \dots, N_d)$ instead of m by N_k as earlier, that is the observations from the last non-complete period are deleted.

We also define recursively

$$\hat{E}_i = (I_N - \sum_{j=1}^{i-1} \hat{E}_j (\hat{E}_j^T \hat{E}_j)^{-1} \hat{E}_j^T) P Y_i^T,$$

for $i=1, \dots, d$, where $P = I_N - X'^T (X'X'^T)^{-1} X'$. Let the p -square reversion matrix R_p be given by

$$R_p = \begin{pmatrix} 0 & 0 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \dots & 0 \\ 1 & 0 & \dots & 0 \end{pmatrix}.$$

THEOREM 4.5. The Yule Walker estimates $\hat{\Psi}^{YW}$ and $\hat{\Theta}^{YW}$ of Ψ and Θ and the deperiodized Yule-Walker estimates $\hat{\Phi}^{dYW}(k)$ and $\hat{\Sigma}^{dYW}(k)$ of $\Phi(k)$ and $\Sigma(k)$, $k=1, \dots, d$, in a PMAR process, based on the estimates C_{00} , $C_{0'}$ and $C_{..}$, are given by

$$\hat{\Psi}^{YW} = C_{0'}, C_{..}^{-1} = XX'^T (X'X'^T)^{-1}$$

$$\hat{\Theta}^{YW} = C_{00} - C_{0'}, C_{..}^{-1} C_{0'} = X P X^T / N$$

$$\hat{\Sigma}^{dYW}(k) = Y_k (P - \sum_{i=1}^{k-1} \hat{E}_i (\hat{E}_i^T \hat{E}_i)^{-1} \hat{E}_i^T) Y_k^T / N = \hat{E}_k^T \hat{E}_k / N$$

for $k=1, \dots, d$, and

$$\hat{\Phi}_j^{dYW}(k) = \begin{cases} (Y_k - \sum_{i=1}^{j-1} \hat{\Phi}_i^{dYW}(k) Y_{k-i}) \hat{E}_{k-j} (\hat{\Sigma}(k-j))^{-1} / N, & j=1, \dots, k-1 \\ (Y_k - \sum_{i=1}^{k-1} \hat{\Phi}_i^{dYW}(k) Y_{k-i}) X'^T (X'X'^T)^{-1} K_j, & j=k, k+1, \dots, pd+k-1 \end{cases}$$

for $k=1, \dots, d$, where

$$K_j = (I_{md} \otimes R_p) (I_m \otimes 1_{pd, pd+k-j}),$$

and $1_{n,j}$ is a vector of dimension n with a one in position j as its only nonzero element. $\square$

Remark 4.1. Although it is possible to decompose $(\hat{\Psi}^{YW})^T$ into $UD^{\frac{1}{2}}L^T$ (cf. Theorem B.2 in Appendix B), such a decomposition is not useful for estimating $\hat{F}$ and $\hat{\Lambda}$ in a decomposition $(\hat{\Psi}^{YW})^T = \hat{F}^T \hat{\Lambda}^T$. This because the choice of $\hat{F}^T$ and $\hat{\Lambda}^T$ from this decomposition is not unique. We can decompose L into $L_1 L_2$ where L_1 and L_2 are both block unit lower triangular matrices. Thus

$$\hat{F}^T = UD^{\frac{1}{2}}L_2^T$$

and

$$\hat{\Lambda}^T = L_1^T$$

are estimates of F^T and Λ^T for any L_2 in the decomposition of L . This ambiguity in the estimates is due to the lack of framework for the structure of F . $\square$

For any block modified Cholesky decomposition LDL^T of W as given in Lemma 3.1, we define for $k=1, \dots, d-1$,

$$L''_k = \begin{pmatrix} L_{k+1,k+1} & \cdots & L_{k+1,d} \\ \vdots & & \vdots \\ L_{dk} & \cdots & L_{dd} \end{pmatrix}, \quad L'_k = \begin{pmatrix} L_{k+1,k} \\ \vdots \\ L_{dk} \end{pmatrix}$$

$$\text{and } D''_k = \bigoplus_{j=k+1}^d D_{jj}.$$

THEOREM 4.6. The Yule-Walker estimates of Ψ and Θ in the deperiodized PMAR process, based on the estimates C_{00} , $C_{\cdot 0}$ and $C_{\cdot \cdot}$ of T_{00} , $T_{\cdot 0}$ and

T_{α} , respectively, are almost sure consistent and Fisher consistent. It holds that

$$(i) \quad \sqrt{N} (\hat{\Psi}^{YW} - \Psi) \xrightarrow{L} N_{md,mdp}(0, \Theta, T_{\alpha}^{-1}) \quad \text{as } N \rightarrow \infty,$$

$$(ii) \quad \sqrt{N} \text{vec}(\hat{\Theta}^{YW} - \Theta) \xrightarrow{L} N_{m^2d^2}(0, 2S_{md} \Theta^{<2>} S_{md}) \quad \text{as } N \rightarrow \infty.$$

and, asymptotically, $\hat{\Psi}^{YW}$ and $\hat{\Theta}^{YW}$ are stochastically independent. Further, for $j=1, \dots, d-1$ we have

$$(iii) \quad \sqrt{N} (\hat{\Lambda}_{\cdot j} - \Lambda_{\cdot j}) \xrightarrow{L} N_{m(d-j),m}(0, \Lambda_{\cdot j}'' (\bigoplus_{k=j+1}^d \Sigma(k)) \Lambda_{\cdot j}''^T, \Sigma(j)^{-1}) \quad \text{as } N \rightarrow \infty,$$

for $j=1, \dots, d-1$ it holds that

$$(iv) \quad \sqrt{N} ((\hat{\Lambda}^{-1})_{\cdot j} - (\Lambda^{-1})_{\cdot j}) \xrightarrow{L} N_{m(d-j),m}(0, \bigoplus_{k=j+1}^d \Sigma(k), \Sigma(j)^{-1}) \quad \text{as } N \rightarrow \infty,$$

and for $j=1, \dots, d$,

$$(v) \quad \sqrt{N} \text{vec}(\hat{\Sigma}^{dYW}(j) - \Sigma(j)) \xrightarrow{L} N_m(0, 2S_m \Sigma(j)^{<2>} S_m) \quad \text{as } N \rightarrow \infty.$$

Furthermore

$$(vi) \quad \sqrt{N} (\hat{F} - F) \xrightarrow{L} N_{md,mdp}(0, \bigoplus_{k=1}^d \Sigma(k), T_{\alpha}^{-1}) \quad \text{as } N \rightarrow \infty.$$

and, asymptotically, $(\hat{\Sigma}^{dYW}(j) : \hat{\Lambda}_{\cdot j}^T)$, $j=1, \dots, d-1$ and $\hat{\Sigma}^{dYW}(d)$ are all stochastically independent.

PROOF. See Appendix C. □

4.2. ESTIMATION IN PMGAR PROCESSES

Let as before $q = \max_{1 \leq j \leq d} (p_j - j)$ and let a sample $y_{-q}, \dots, y_0, y_1, \dots, y_n$ of a PMGAR(P_d^T) process be available. Further, let

$$L = L(\phi_i(k), \Sigma(k), k=1, \dots, d, i=1, \dots, p_k)$$

denote the likelihood function of the parameters given the data. By conditioning on $y_{-q}, \dots, y_0$ we have

$$L = f_{(1,n)} | (-q, 0) (y_1, \dots, y_n | y_{-q}, \dots, y_0) f_{(-q, 0)} (y_{-q}, \dots, y_0)$$

where $f_{(1,n)|(-q,0)}$ denotes the simultaneous density of $y_1, \dots, y_n$ given $y_{-q}, \dots, y_0$, and $f_{(-q,0)}$ denotes the simultaneous density of $y_{-q}, \dots, y_0$.

Due to the assumption of normality and mutually independence of the e_k 's, we infer that $\text{vec}(y_{-q} : \dots : y_0)$ has a multivariate normal distribution and thus

$$f_{(-q,0)}(y_{-q}, \dots, y_0) = (2\pi)^{-\frac{1}{2}(q+1)m} |C_{-q,0}|^{-\frac{1}{2}} \exp(-\frac{1}{2}(\text{vec}(y_{-q} : \dots : y_0))^T C_{-q,0}^{-1} \text{vec}(y_{-q} : \dots : y_0))$$

where $C_{-q,0}$ denotes the dispersion matrix of $\text{vec}(y_{-q} : \dots : y_0)$.

For determining the density $f_{(1,n)|(-q,0)}(y_1, \dots, y_n | y_{-q}, \dots, y_0)$, consider for fixed $y_{-q}, \dots, y_0$ the mapping

$$g: e_1, \dots, e_n \rightarrow y_1, \dots, y_n$$

defined by (2.1). The Jacobian of this mapping is equal to one, and thus, again due to the normality and mutually independence of $e_1, \dots, e_n$,

$$(4.10) \quad f_{(1,n)|(-q,0)}(y_1, \dots, y_n | y_{-q}, \dots, y_0) = (2\pi)^{-\frac{1}{2}mn} \prod_{i=1}^n |\Sigma(i)|^{-\frac{1}{2}} \exp(-\frac{1}{2} \sum_{i=1}^n (y_i - \sum_{j=1}^{p_i} \phi_j(i) y_{i-j})^T \Sigma(i)^{-1} (y_i - \sum_{j=1}^{p_i} \phi_j(i) y_{i-j}))$$

$$= (2\pi)^{-\frac{1}{2}mn} \prod_{k=1}^d |\Sigma(k)|^{-N_k/2} \exp(-\frac{1}{2} \sum_{k=1}^d \text{tr}((Y_k - \Phi(k)Y'_k)^T \Sigma(k)^{-1} (Y_k - \Phi(k)Y'_k)))$$

where in the last equality we have taken into account the periodicity in k of $\Phi(k)$ and $\Sigma(k)$.

We can thus write the likelihood function

$$L = (2\pi)^{-\frac{1}{2}m(n+q+1)} |C_{-q,0}|^{-\frac{1}{2}} \left(\prod_{k=1}^d |\Sigma(k)|^{N_k} \right) \exp(-\frac{1}{2} \sum_{j=0}^d q(j))$$

where

$$q(k) = \begin{cases} (\text{vec}(y_{-q} : \dots : y_0))^T C_{-q,0}^{-1} \text{vec}(y_{-q} : \dots : y_0), & k=0 \\ \text{tr}((Y_k - \Phi(k)Y'_k)^T \Sigma(k)^{-1} (Y_k - \Phi(k)Y'_k)), & k=1, \dots, d. \end{cases}$$

4.2.1. Maximum conditional likelihood estimates

For any matrices X and Y the outer product $X \square Y$ denotes the matrix $\text{vec}(Y)(\text{vec}(X))^T$. The matrix $d_X \square Y$ with $d_X = (\partial/\partial x_{ij})$ is a matrix of partial derivatives of the elements of Y with respect to the elements of X . For an arbitrary parameter matrix Ξ we define the Fisher information matrix $I(\Xi, \Xi)$, as the arrangement

$$I(\Xi, \Xi) = - E(d_{\Xi} \square (d_{\Xi} \square \ln f)).$$

where $\ln$ is the natural logarithm.

This implies that for $\Xi = (\phi(1): \dots : \phi(d): \Sigma(1): \dots : \Sigma(d))$ the information matrix can be written

$$I(\Xi, \Xi) = \begin{pmatrix} I(\phi(1), \phi(1)) & \dots & I(\phi(1), \phi(d)): I(\phi(1), \Sigma(1)) & \dots & I(\phi(1), \Sigma(d)) \\ \vdots & & \vdots & & \vdots \\ I(\phi(d), \phi(1)) & \dots & I(\phi(d), \phi(d)): I(\phi(d), \Sigma(1)) & \dots & I(\phi(d), \Sigma(d)) \\ I(\Sigma(1), \phi(1)) & \dots & I(\Sigma(1), \phi(d)): I(\Sigma(1), \Sigma(1)) & \dots & I(\Sigma(1), \Sigma(d)) \\ \vdots & & \vdots & & \vdots \\ I(\Sigma(d), \phi(1)) & \dots & I(\Sigma(d), \phi(d)): I(\Sigma(d), \Sigma(1)) & \dots & I(\Sigma(d), \Sigma(d)) \end{pmatrix}$$

where

$$I(\phi(i), \phi(j)) = - E(d_{\phi(i)} \square (d_{\phi(j)} \square \ln f))$$

$$I(\Sigma(i), \phi(j)) = - E(d_{\Sigma(i)} \square (d_{\phi(j)} \square \ln f))$$

$$-I(\phi(i), \Sigma(j)) = - E(d_{\phi(i)} \square (d_{\Sigma(j)} \square \ln f)) = I(\Sigma(j), \phi(i))^T$$

and

$$I(\Sigma(i), \Sigma(j)) = - E(d_{\Sigma(i)} \square (d_{\Sigma(j)} \square \ln f))$$

for $i, j=1, \dots, d$, are the Fisher information matrices and cross information matrices of the parameters respectively. Each of the matrices $I(\phi(i), \phi(j))$ can be partitioned further according to the partition $(\phi_1(k): \dots : \phi_{p_k}(k))$ of $\phi(k)$. The following result for the PGMAR process shows that $I(\Xi, \Xi)$ is blockdiagonal so that $I(\phi(i), \phi(j)) = 0$ and $I(\Sigma(i), \Sigma(j)) = 0$ for $i \neq j$ and $I(\Sigma(i), \phi(j)) = 0$ for all i and j .

THEOREM 4.7. The Fisher information of $\Xi = (\phi(1): \dots : \phi(d): \Sigma(1): \dots : \Sigma(d))$

contained in $y_1, \dots, y_n$ given $y_{-q}, \dots, y_0$, is block diagonal and given by

$$I(\Xi, \Xi) = \bigoplus_{k=1}^d \left(N_k (\Sigma(k)^{-1} \otimes \Gamma_{\pi}(k)) \right) \oplus \left(\bigoplus_{k=1}^d \frac{1}{2} N_k S_m (\Sigma(k)^{-1} \otimes \Sigma(k)^{-1}) S_m \right)$$

PROOF. See Appendix D. □

The result of Theorem 4.7 for PMGAR processes is similar to the result of Pagano (1978) for the (univariate) PGAR processes although Pagano considers the Fisher information matrix of the parameters contained in a unconditional data set.

DEFINITION 4.4. Values $\hat{\Phi}^{CL}(k)$ and $\hat{\Sigma}^{CL}(k)$ of $\Phi(k)$ and $\Sigma(k)$ which maximize the conditional likelihood given by (4.10) are called maximum conditional likelihood (m.c.l.) estimates of $\Phi(k)$ and $\Sigma(k)$. □

Recall the definition of $Q_{ij}(A)$ below (4.1). In this notation we have the following result.

THEOREM 4.8. The maximum conditional likelihood estimates $\hat{\Phi}^{CL}(k)$ and $\hat{\Sigma}^{CL}(k)$, of $\Phi(k)$ and $\Sigma(k)$, $k=1, \dots, d$, in a $\text{PMGAR}(\mathbf{p}_d^T)$ process are given by

$$\hat{\Phi}^{CL}(k) = \arg(\min_{\Phi(k)} (|Q_{kk}(\Phi(k))|)), \quad k=1, \dots, d$$

and

$$\hat{\Sigma}^{CL}(k) = N_k^{-1} \min_{\Phi(k)} (Q_{kk}(\Phi(k))) = N_k^{-1} Q_{kk}(\hat{\Phi}^{CL}(k)), \quad k=1, \dots, d.$$

PROOF. The estimates $\hat{\Phi}^{CL}(k)$ and $\hat{\Sigma}^{CL}(k)$, $k=1, \dots, d$ are the values of $\Phi(k)$ and $\Sigma(k)$, $k=1, \dots, d$, that minimize

$$\prod_{k=1}^d |\Sigma(k)|^{-\frac{1}{2} N_k} \exp(-\frac{1}{2} \sum_{k=1}^d q(k)).$$

Since $q(k)$ just depends on $\Phi(k)$ and $\Sigma(k)$, the expression above is maximized if for each k , $|\Sigma(k)|^{-\frac{1}{2} N_k} \exp(-\frac{1}{2} q(k))$ is maximized. For fixed $\Phi(k)$ it follows from Lemma G.1 in Appendix G that the latter expression has a maximum when

$$(4.11) \quad \Sigma(k) = (Y_k - \Phi(k)Y'_k)(Y_k - \Phi(k)Y'_k)^T / N_k$$

which by definition is equal to $Q_{kk}(\Phi(k))/N_k$. The maximum value, for fixed $\Phi(k)$, is equal to

$$\frac{mN_k/2}{N_k} |Q_{kk}(\Phi(k))|^{-1/2} \exp(-1/2 mN_k)$$

and thus $\hat{\Phi}^{CL}(k)$ is the value of $\Phi(k)$ which minimizes $|Q_{kk}(\Phi(k))|$.

Hence the theorem is proved. $\square$

The preceding theorem does not give explicit expressions of the estimates in terms of the observations, but merely states the nature of the estimates in terms of minimizing a loss function. It is seen from the expressions in Theorem 4.8 that $\hat{\Phi}^{CL}(k)$ and $\hat{\Sigma}^{CL}(k)$ are least generalized squares estimates of $\Phi(k)$ and $\Sigma(k)$ respectively.

The result of the next theorem follows from that an LG estimate is also a LGS estimate and from (4.11).

THEOREM 4.9. The m.c.l. estimates $\hat{\Phi}^{CL}(k)$ and $\hat{\Sigma}^{CL}(k)$ of $\Phi(k)$ and $\Sigma(k)$, $k=1, \dots, d$, in a PMGAR(p_d^T) process are given by

$$\hat{\Phi}^{CL}(k) = Y_k Y'_k{}^T (Y'_k Y'_k{}^T)^{-1}, \quad k=1, \dots, d$$

and

$$\hat{\Sigma}^{CL}(k) = Y_k (I_{N_k} - Y'_k{}^T (Y'_k Y'_k{}^T)^{-1} Y'_k) Y_k^T / N_k, \quad k=1, \dots, d. \quad \square$$

From the expressions of Theorem 4.9 we see that the m.c.l. estimates are also equal to the Yule-Walker estimates based on $\{G_{ij}(k)\}_{i,j}, k=1, \dots, d$.

4.2.2. Approximate maximum likelihood estimates

The maximum conditional likelihood estimates given in the last section do not take into account the initial $q+1$ observed values of the process. This is of minor importance if n is large but may have an effect for moderate n . In order to take into account these observations when maximizing the like-

likelihood, we need an expression for $C_{-1,0}$ in terms of $\Phi(k)$, $k=1, \dots, d$. To get an idea how to do this we first consider a problem similar to this in the deperiodized process.

Suppose that a realization $x_{-p+1}, \dots, x_0, x_1, \dots, x_N$ of the deperiodized process is available. Because of the normality of $\{y_k\}$, the likelihood of the parameters in the deperiodized process, given the data, is

$$\begin{aligned} L_0 &= L_0(\Psi_1, \dots, \Psi_p, \Theta) = (2\pi)^{-\frac{1}{2}md(N+p)} |W_{N+p}|^{-\frac{1}{2}} \\ &\times \exp(-\frac{1}{2}\text{tr}(W_{N+p}^{-1}((x_{-p+1} \dots x_0 : x_1 \dots x_N) \square (x_{-p+1} \dots x_0 : x_1 \dots x_N))) \\ &= (2\pi)^{-\frac{1}{2}md(N+p)} |W_p|^{-\frac{1}{2}} |\Theta|^{-\frac{1}{2}N} \exp(-\frac{1}{2} \sum_{i=1}^N (x_i - \Psi x_i')^T \Theta^{-1} (x_i - \Psi x_i')) \\ &\times \exp(-\frac{1}{2}\text{tr}(W_p^{-1}((x_{-p+1} \dots x_0) \square (x_{-p+1} \dots x_0)))) \end{aligned}$$

where for general n , $W_n = E((x_1 \dots x_n) \square (x_1 \dots x_n))$ is the covariance matrix of n successive samples, i.e. for example of $\text{vec}(x_1 \dots x_n)$.

Given any $n > p$ we can express W_n^{-1} in terms of the matrices $\{\Psi_i\}$ and the inverse of the covariance matrix W_p of $\text{vec}(x_1 \dots x_p)$. An application of a result of Anděl (1971, Theorem 6) yields that the (i, j) 'th block W_n^{ij} of the inverse of W_n is given by the formulas

$$(4.12) \quad W_n^{ij} = \begin{cases} W_p^{ij} + \sum_{k=p+1}^n B_{k-i} \Theta^{-1} B_{k-j}^T & \text{if } 1 \leq i, j \leq p \\ \sum_{k=p+1}^n B_{k-i} \Theta^{-1} B_{k-j}^T & \text{if } i > p \text{ or } j > p \end{cases}$$

where $B_j^T = -\Psi_j$, $j=1, \dots, p$, $B_j = 0$ if $j > p$ or $j < 0$ and $B_0 = I_{md}$.

In the one dimensional case, this relation for $n=p+1$ and the knowledge that the covariance matrix is doubly symmetric (symmetric and persymmetric; symmetric in both main diagonals) are used to obtain an explicit formula for W_p^{-1} . In the multivariate, case the covariance matrix is not generally persymmetric "in the blocks" as noted by Anděl, and the same procedure can not be applied to obtain an expression for W_p^{-1} .

When the time axis can be reverted without changing the properties of the time series, it is possible to obtain an explicit expression for W_p^{-1} and

hence also, from (4.12), for W_n^{-1} . Following Siddiqui (1958) and Durbin (1959) for the univariate case, we have the two following expressions for the density of $\text{vec}(x_1: \dots: x_n: x_{n+1}: \dots: x_{2n})$:

$$(4.13) \quad c_1 c_2 \exp\left(-\frac{1}{2} \left(\sum_{i=1}^n (x_{n+i} - \sum_{j=1}^n \psi_j x_{n+i-j})^T \Theta^{-1} (x_{n+i} - \sum_{j=1}^n \psi_j x_{n+i-j}) \right.\right. \\ \left.\left. \exp\left(-\frac{1}{2} (\text{vec}(x_1: \dots: x_n))^T W_n^{-1} \text{vec}(x_1: \dots: x_n)\right) \right)$$

and

$$(4.14) \quad c_1 c_2 \exp\left(-\frac{1}{2} \left(\sum_{i=1}^n (x_i - \sum_{j=1}^n \psi_j x_{i+j})^T \Theta^{-1} (x_i - \sum_{j=1}^n \psi_j x_{i+j}) \right.\right. \\ \left.\left. \exp\left(-\frac{1}{2} (\text{vec}(x_{n+1}: \dots: x_{2n}))^T W_n^{-1} \text{vec}(x_{n+1}: \dots: x_{2n})\right) \right).$$

for $n \geq p$ where $\psi_j = 0$ when $j > p$. Here the first exponential terms in (4.13) and (4.14) represent the densities conditional on $\text{vec}(x_1: \dots: x_n)$ and $\text{vec}(x_{n+1}: \dots: x_{2n})$ respectively, and the second exponential terms represent the densities of $\text{vec}(x_1: \dots: x_n)$ and $\text{vec}(x_{n+1}: \dots: x_{2n})$, respectively. The constants c_1 and c_2 are normalising constants. In (4.14) the regression is on future observations, or equivalently, the time axis is reverted, which by assumption should have no effect on the time series. Let $B_j^T = -\psi_j$, for $j=1, \dots, p$, $B_j = 0$, if $j > p$ or $j < 0$, $B_0 = I_{md}$ and define

$$A_{ij} = \begin{pmatrix} B_i \Theta^{-1} B_i^T & B_i \Theta^{-1} B_{i+1}^T & \dots & B_i \Theta^{-1} B_j^T \\ B_{i+1} \Theta^{-1} B_i^T & B_{i+1} \Theta^{-1} B_{i+1}^T & \dots & B_{i+1} \Theta^{-1} B_j^T \\ \vdots & \vdots & & \vdots \\ B_j \Theta^{-1} B_i^T & B_j \Theta^{-1} B_{i+1}^T & \dots & B_j \Theta^{-1} B_j^T \end{pmatrix} \quad \text{for } i \leq j$$

and

$$A_{ij} = \begin{pmatrix} B_i \Theta^{-1} B_i^T & B_i \Theta^{-1} B_{i-1}^T & \dots & B_i \Theta^{-1} B_j^T \\ B_{i-1} \Theta^{-1} B_i^T & B_{i-1} \Theta^{-1} B_{i-1}^T & \dots & B_{i-1} \Theta^{-1} B_j^T \\ \vdots & \vdots & & \vdots \\ B_j \Theta^{-1} B_i^T & B_j \Theta^{-1} B_{i-1}^T & \dots & B_j \Theta^{-1} B_j^T \end{pmatrix} \quad \text{for } i > j.$$

By equating (4.13) to (4.14) and regarding terms containing $x_1, \dots, x_n$, we

obtain

$$\begin{aligned} & (\text{vec}(x_1 : \dots : x_n))^T (W_n^{-1} + \sum_{k=1}^n A_{n+k-1, k}) \text{vec}(x_1 : \dots : x_n) \\ & = (\text{vec}(x_1 : \dots : x_n))^T \left(\sum_{k=1}^n A_{1-k, n-k} \right) \text{vec}(x_1 : \dots : x_n) \end{aligned}$$

and thus

$$W_n^{-1} = \sum_{k=1}^n (A_{1-k, n-k} - A_{n+k-1, k}).$$

Similarly, by regarding terms containing $x_{n+1}, \dots, x_{2n}$ we obtain

$$W_n^{-1} = \sum_{k=1}^n (A_{n-k, 1-k} - A_{k, n+k-1}).$$

The (i, j) 'th block of order md by md of W_n^{-1} is thus given by the expressions

$$(4.15) \quad W_n^{ij} = \sum_{k=1}^n (B_{i-k} \Theta^{-1} B_{j-k}^T - B_{n+k-i} \Theta^{-1} B_{n+k-j}^T)$$

and

$$(4.16) \quad W_n^{ij} = \sum_{k=1}^n (B_{n-k+1-i} \Theta^{-1} B_{n-k+1-j}^T - B_{k+i-1} \Theta^{-1} B_{k+j-1}^T).$$

Here W_n^{ij} is a null matrix if $|i-j| > p$ so that W_n^{-1} is a block band matrix with block bandwidth $2p+1$. A sufficient condition for the equality of (4.15) and (4.16) is

$$B_i \Theta^{-1} B_j^T = B_j \Theta^{-1} B_i^T$$

i.e. that $B_i \Theta^{-1} B_j^T$ is symmetric for $i, j=1, \dots, p$. For then, for each k , the term $B_{i-k} \Theta^{-1} B_{j-k}^T - B_{n+k-i} \Theta^{-1} B_{n+k-j}^T$ in (4.15) corresponds to a specific term $B_{n-k'+1-i} \Theta^{-1} B_{n-k'+1-j}^T - B_{k'+i-1} \Theta^{-1} B_{k'+j-1}^T$ in (4.16) where $k'=k+n+1-i-j$.

Expressions for the univariate case corresponding to (4.15) and (4.16) were given by Siddiqui (1958), Durbin (1959) and recently also by Anderson and Mentz (1980).

A simple way to express W_p^{-1} is to introduce two block unit lower triangular matrices $M_{0:p}$ and $M_{p:0}$ defined by

$$M_{0:p} = \begin{pmatrix} I & 0 & \dots & 0 \\ B_1 & I & \dots & 0 \\ B_2 & B_1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ B_{p-1} & B_{p-2} & \dots & I \end{pmatrix}$$

and

$$M_{p:0} = \begin{pmatrix} B_p & 0 & \dots & 0 \\ B_{p-1} & B_p & \dots & 0 \\ B_{p-2} & B_{p-1} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ B_1 & B_2 & \dots & B_p \end{pmatrix}$$

then

$$W_p^{-1} = M_{0:p}(\Theta^{-1} \otimes I_p)M_{0:p}^T - M_{p:0}(\Theta^{-1} \otimes I_p)M_{p:0}^T.$$

The corresponding expression for univariate autoregressions was proposed by Pagano (1973).

By using this representation of W_p^{-1} it is seen that

$$\begin{aligned} \sum_{i=1}^N (x_i - \sum_{j=1}^p \psi_j x_{i-j})^T \Theta^{-1} (x_i - \sum_{j=1}^p \psi_j x_{i-j}) &+ \text{tr}(W_p^{-1}((x_{-p+1} : \dots : x_0) \square (x_{-p+1} : \dots : x_0))) \\ &= \text{tr}(\sum_{i=0}^p \sum_{j=0}^p B_i \Theta^{-1} B_j^T D_{ji}) \end{aligned}$$

where

$$(4.17) \quad D_{ji} = \sum_{u=i+j+1-p}^N x_{u-j} x_{u-i}^T.$$

Now, differentiation of the log-likelihood gives

$$(4.18) \quad d_{B_k} \square \ln L_0 = \frac{1}{2} M_k + \frac{1}{2} d_{B_k} \square \text{tr}(\sum_{i=0}^p \sum_{j=0}^p B_i \Theta^{-1} B_j^T D_{ji})$$

where $M_k = d_{B_k} \square |W_p|$. We have

$$(4.19) \quad \frac{1}{2} d_{B_k} \square \text{tr}(\sum_{i=0}^p \sum_{j=0}^p B_i \Theta^{-1} B_j^T D_{ji}) = \text{vec}(\sum_{i=0}^p \tilde{D}_{ki} B_i \Theta^{-1})$$

where $\tilde{D}_{ki} = \frac{1}{2}(D_{ki} + D_{ik})$, and thus by equating the expressions for $k=1, \dots, p$ to zero and solving for $\Psi = (\psi_1 : \dots : \psi_p)$ or equivalently for $B_1, \dots, B_p$, we

have

$$\hat{\Psi} = (\hat{\Psi}_1 : \dots : \hat{\Psi}_p) = (\tilde{D}_{01} : \dots : \tilde{D}_{0p}) \tilde{D}_{\Pi}^{-1}$$

where

$$\tilde{D}_{\Pi} = \begin{pmatrix} \tilde{D}_{11} & \dots & \tilde{D}_{1p} \\ \vdots & & \vdots \\ \tilde{D}_{p1} & \dots & \tilde{D}_{pp} \end{pmatrix}.$$

By taking expectations in (4.18) and using (4.19) we get

$$(4.20) \quad \frac{1}{2} M_k + \text{vec} \left(\sum_{i=0}^p (N+p-i-k) T_{ki} B_i \Theta^{-1} \right).$$

We have that $E(d_{B_k} \square \ln L_0) = 0$ and since $\sum_{i=0}^p T_{ki} B_i = 0$, for $k > 0$ (Theorem 4.4 (i)), it follows by post-multiplication with $(N+p)\Theta^{-1}$ and subtraction from (4.20) that the expectation of

$$\frac{1}{2} M_k - \text{vec} \left(\sum_{i=0}^p (i+k) T_{ki} B_i \Theta^{-1} \right)$$

is zero. We use $\tilde{D}_{ki} / (N+p-i-k)$ as an unbiased estimate of T_{ki} , and a natural estimate of M_k is then

$$\hat{M}_k = 2 \text{vec} \left(\sum_{i=0}^p (i+k) \tilde{D}_{ki} B_i \Theta^{-1} / (N+p-i-k) \right).$$

Substitution of this estimate into (4.18) yields

$$d_{B_k} \square \ln L_0 \cong (N+p) \text{vec} \left(\sum_{i=0}^p \tilde{D}_{ki} B_i \Theta^{-1} / (N+p-i-k) \right)$$

for $k=0,1,\dots,p$. By equating these expressions to zero and solving the equations, we get

$$\tilde{\Psi} = (\tilde{D}_{01} : \dots : \tilde{D}_{0p}) \tilde{D}_{\Pi}^{-1}$$

where $\tilde{D}_{ij} = \tilde{D}_{ij} / (N+p-i-j)$ and $\tilde{D}_{\Pi}$ is obtained from $\tilde{D}_{\Pi}$ by replacing $\tilde{D}_{ij}$ by $\tilde{D}_{ij}$.

Returning to the nonperiodized process, we can now propose the estimate $\hat{\Phi}(k)$ of $\Phi(k)$ given by

$$\hat{\Phi}(k) = \tilde{G}_0(k) \tilde{G}_{\Pi}(k)^{-1}$$

where $\tilde{\tilde{G}}_0(k) = (\tilde{\tilde{G}}_{01}(k) : \dots : \tilde{\tilde{G}}_{0p_k}(k))$ and

$$\tilde{\tilde{G}}_{ii}(k) = \begin{pmatrix} \tilde{\tilde{G}}_{11}(k) & \dots & \tilde{\tilde{G}}_{1p_k}(k) \\ \vdots & & \vdots \\ \tilde{\tilde{G}}_{p_k 1}(k) & \dots & \tilde{\tilde{G}}_{p_k p_k}(k) \end{pmatrix}.$$

Here we define $\tilde{\tilde{G}}_{ij}(k)$ in a way suitable for the periodic case and analogous to (4.17),

$$\tilde{\tilde{G}}_{ij}(k) = \sum_{u=i+j+1-p}^{N_k} y_{k-i+(u-1)d} y_{k-j+(u-1)d}^T / (N_k + p - i - j)$$

for $i, j=1, \dots, p_k$.

5. PARAMETER ESTIMATION IN SOME REDUCED MODELS

There are two special cases of the general model which are of interest. One of these concerns the case when either the innovation process covariances or the regression matrices are independent of time. The other interesting special case of the general model concerns a reduction of the interrelation structure between the subprocesses and will be divided into the case of no interrelation between the corresponding innovation subprocesses and the case when there is no interrelation between the subprocesses through the regression matrices.

5.1. TIME HOMOGENEITY

If the regression matrices $\Phi(k)$, $k=1, \dots, d$, in (2.1) are independent of time, the model is referred to as a homogeneous regression PMAR process and is denoted by P_R^{MAR} . If the innovation covariances $\Sigma(k)$, $k=1, \dots, d$, in (2.1) are independent of time, the model is referred to as a homogeneous innovation PMAR process and is denoted by P_I^{MAR} .

5.1.1. Estimation in the P_R^{MAR} process

In the P_R^{MAR} process we have $p_1 = \dots = p_d = p'$ say. We write ϕ_i , $i=1, \dots, p'$ for the regression matrices and define Φ by $\Phi = (\phi_1 : \dots : \phi_{p'})$. We introduce

$$y_k'' = \text{vec}(y_{k-1} : \dots : y_{k-p'})$$

and

$$Y_k'' = (y_k'' : y_{k+d}'' : \dots : y_{k+(N_k-1)d}'')$$

where the extra prime in comparison to y_k' and Y_k' indicates the situation at hand.

DEFINITION 5.1. Let F be a parameter space for Φ . A random variable $\hat{\Phi}_R^{\text{LG}}$

belonging to F a.s. such that the relations

$$\begin{aligned} (Y_1 - \hat{\phi}_R^{LG} Y_1'')(Y_1 - \hat{\phi}_R^{LG} Y_1'')^T &\leq (Y_1 - \phi Y_1'')(Y_1 - \phi Y_1'')^T \\ &\vdots \\ (Y_d - \hat{\phi}_R^{LG} Y_d'')(Y_d - \hat{\phi}_R^{LG} Y_d'')^T &\leq (Y_d - \phi Y_d'')(Y_d - \phi Y_d'')^T \end{aligned}$$

all hold a.s. for all ϕ in F , is called a least Gramian estimator of ϕ in the P_{RMAR} process. A class of least Gramian estimators $\{\hat{\Sigma}_R^{LG}(k), k=1, \dots, d\}$ of $\{\Sigma(k), k=1, \dots, d\}$ in the P_{RMAR} process is defined by

$$\hat{\Sigma}_R^{LG}(k) = (Y_k - \hat{\phi}_R^{LG} Y_k'')(Y_k - \hat{\phi}_R^{LG} Y_k'')^T / f_k^R$$

where f_k^R is a suitably chosen number.

THEOREM 5.1. Let $M((Y_1 Y_1'^T : \dots : Y_d Y_d'^T)^T) \subset M((Y_1'' Y_1''^T : \dots : Y_d'' Y_d''^T)^T)$ a.s. and let F be the range space of $(Y_1 Y_1'^T : \dots : Y_d Y_d'^T)$, then there exists an LG estimator $\hat{\phi}_R^{LG}$ of $\phi \in F$ in the homogeneous regression PMAR model. An estimator can be taken to be

$$(5.1) \quad \hat{\phi}_R^{LG} = (Y_1 Y_1'^T : \dots : Y_d Y_d'^T) (Y_1'' Y_1''^T : \dots : Y_d'' Y_d''^T)^{-}$$

for an arbitrary g-inverse $(Y_1'' Y_1''^T : \dots : Y_d'' Y_d''^T)^{-}$ of $(Y_1'' Y_1''^T : \dots : Y_d'' Y_d''^T)$.

PROOF. Write $\phi = \hat{\phi}_R^{LG} - \Delta$, where $\hat{\phi}_R^{LG}$ is given by (5.1). The assumption implies that $(Y_1 Y_1'^T : \dots : Y_d Y_d'^T) (Y_1'' Y_1''^T : \dots : Y_d'' Y_d''^T)^{-} (Y_1'' Y_1''^T : \dots : Y_d'' Y_d''^T) = (Y_1 Y_1'^T : \dots : Y_d Y_d'^T)$ and thus $(Y_k - \hat{\phi}_R^{LG} Y_k'')(\Delta Y_k'')^T = 0$ for $k=1, \dots, d$.

Hence we have that

$$\begin{aligned} (Y_k - \phi Y_k'')(Y_k - \phi Y_k'')^T &= (Y_k - \hat{\phi}_R^{LG} Y_k'')(Y_k - \hat{\phi}_R^{LG} Y_k'')^T + \Delta Y_k'' Y_k''^T \Delta^T \\ &\geq (Y_k - \hat{\phi}_R^{LG} Y_k'')(Y_k - \hat{\phi}_R^{LG} Y_k'')^T \\ &= (Y_k - \hat{\phi}_R^{LG} Y_k'') Y_k^T = Y_k (Y_k - \hat{\phi}_R^{LG} Y_k'')^T \end{aligned}$$

for $k=1, \dots, d$, which proves the theorem.

5.1.2. Estimation in the P_{IMAR} process

If the innovation covariances $\Sigma(k), k=1, \dots, d$ are independent of time we

write Σ for $\Sigma(k)$. The interpretation of the definition of least Gramian estimates for this situation is given in the following definition.

DEFINITION 5.2. Let $F(k)$ be a parameter space for $\phi(k)$. A set $\{\hat{\phi}_I^{LG}(k) \in F(k) \text{ a.s., } k=1, \dots, d\}$ of random variables such that the relations

$$(5.2) \quad \begin{array}{ccc} (Y_1 - \hat{\phi}_I^{LG}(1)Y'_1)(Y_1 - \hat{\phi}_I^{LG}(1)Y'_1)^T & \Xi & (Y_1 - \phi(1)Y'_1)(Y_1 - \phi(1)Y'_1)^T \\ \vdots & & \vdots \\ (Y_d - \hat{\phi}_I^{LG}(d)Y'_d)(Y_d - \hat{\phi}_I^{LG}(d)Y'_d)^T & \Xi & (Y_d - \phi(d)Y'_d)(Y_d - \phi(d)Y'_d)^T \end{array}$$

all holds a.s. for all $\phi(k) \in F(k)$, is called a set of least Gramian estimators of $\{\phi(k), k=1, \dots, d\}$ in the P_I MAR process.

A class of least Gramian estimator $\hat{\Sigma}_I^{LG}$ of Σ in the P_I MAR process is defined by

$$\hat{\Sigma}_I^{LG} = \sum_{k=1}^d (Y_k - \hat{\phi}_I^{LG}(k)Y'_k)(Y_k - \hat{\phi}_I^{LG}(k)Y'_k)^T / f^I$$

where f^I is a suitably chosen number.

We have the following results for this case.

THEOREM 5.2. There exists a set $\{\hat{\phi}_I^{LG}(k) \in F(k) \text{ a.s., } k=1, \dots, d\}$ of LG estimators in the homogeneous innovation PMAR model. An estimator can be taken to be

$$\hat{\phi}_I^{LG}(k) = Y_k Y'_k{}^T (Y'_k Y'_k{}^T)^{-}, \quad k=1, \dots, d$$

and a LG estimator $\hat{\Sigma}_I^{LG}$ of Σ is given by

$$(5.3) \quad \hat{\Sigma}_I^{LG} = (Y_1 : \dots : Y_d) (I_n - (\bigoplus_{k=1}^d Y'_k Y'_k{}^T)^{-} Y'_k) (Y_1 : \dots : Y_d)^T / (n-r^I).$$

where $r^I = \sum_{k=1}^d r_k^o$ and $r_k^o = \text{tr}(Y'_k Y'_k{}^T)^{-} Y'_k$.

PROOF. It immediately follows from Theorem 4.1 that the quantities

$\hat{\phi}_I^{LG}(k)$, $k=1, \dots, d$, given by (4.3) satisfy the relations (5.2). The LG

estimator $\hat{\Sigma}_I^{LG}$ is therefore given by

$$\hat{\Sigma}_I^{LG} = \sum_{k=1}^d (Y_k - Y_k Y'_k{}^T (Y'_k Y'_k{}^T)^{-} Y'_k) (Y_k - Y_k Y'_k{}^T (Y'_k Y'_k{}^T)^{-} Y'_k)^T / f^I$$

$$= \sum_{k=1}^d f_k \cdot \hat{\Sigma}^{LG}(k) / f^I.$$

A possible choice of f^I is $f^I = \sum_{k=1}^d f_k = \sum_{k=1}^d (N_k - r_k^0)$ which result in that the least Gramian estimator is a weighted average of the individual covariance matrices. The expression can be rewritten in the form given in the theorem.

5.1.3. Estimation in the $P_{RI}MAR$ process

In case neither the regression parameters nor the innovation covariances depend on time, the model (2.1) reduces to the ordinary MAR process. The LG estimator can be obtained from Theorems 4.1, 5.1, 5.2 or by setting $d=1$. However, for testing purposes it is convenient to have the definition of LG estimators and expressions for these given in terms of an arbitrary predetermined positive integer d , and in the data matrices for the PMAR process of period d .

DEFINITION 5.3. Let F be a parameter space for Φ . A random variable

$\hat{\Phi}_{RI}^{LG}$ belonging to F a.s. such that the relation

$$\sum_{k=1}^d (Y_k - \hat{\Phi}_{RI}^{LG} Y_k'') (Y_k - \hat{\Phi}_{RI}^{LG} Y_k'')^T \leq \sum_{k=1}^d (Y_k - \Phi Y_k'') (Y_k - \Phi Y_k'')^T$$

holds a.s. for all $\Phi \in F$, is called a least Gramian estimator of Φ in the MAR model. A class of least Gramian estimators $\hat{\Sigma}_{RI}^{LG}$ of Σ is defined by

$$\hat{\Sigma}_{RI}^{LG} = \sum_{k=1}^d (Y_k - \hat{\Phi}_{RI}^{LG} Y_k'') (Y_k - \hat{\Phi}_{RI}^{LG} Y_k'')^T / f^{RI}$$

where f^{RI} is a suitably chosen number.

THEOREM 5.3. Let F be the range space of the aperiodic ($d=1$) PMAR process given by (2.1). Then there exists an LG estimator $\hat{\Phi}_{RI}^{LG}$ of $\Phi \in F$. An estimator can be taken to be

$$\hat{\Phi}_{RI}^{LG} = (Y_1 : \dots : Y_d) (Y_1'' : \dots : Y_d'')^T ((Y_1'' : \dots : Y_d'') (Y_1'' : \dots : Y_d'')^T)^{-1}$$

$$= \left(\sum_{k=1}^d \hat{\Phi}^{LG}(k) (Y_k'' Y_k''^T) \right) \left(\sum_{k=1}^d (Y_k'' Y_k''^T) \right)^{-}$$

and an LG estimator $\hat{\Sigma}^{LG}$ of Φ is given by

$$(5.4) \quad \hat{\Sigma}_{RI}^{LG} = \sum_{k=1}^d (Y_k - \hat{\Phi}_{RI}^{LG} Y_k'') (Y_k - \hat{\Phi}_{RI}^{LG} Y_k'')^T / f^{RI}$$

where $f^{RI} = \sum_{k=1}^d (N_k - r_k^0)$.

5.2. REDUCED CORRELATION STRUCTURE

5.2.1. Estimation in the PSMAR process

When considering a simple correlated process it is convenient to partition

$\Phi_i(k)$ into

$$\Phi_i(k) = \begin{pmatrix} \Phi_{i,11}(k) & \dots & \Phi_{i,1s}(k) \\ \vdots & & \vdots \\ \Phi_{i,s1}(k) & \dots & \Phi_{i,ss}(k) \end{pmatrix}$$

where $\Phi_{i,uv}(k)$ for $u,v=1,\dots,s$ is an m_u by m_v matrix. In the simple correlated case we have $\Phi_{i,uv}(k) = 0$ for $u \neq v$ so that

$$\Phi_i(k) = \bigoplus_{j=1}^s \Phi_{i,jj}(k) = B_{\underline{m}_s}(\Phi_i(k)),$$

where $B_{\underline{m}_s}$ is an operator such that $B_{\underline{m}_s}(A)$ is the block-diagonal matrix with the same diagonal blocks, of size m_i -square, $i=1,\dots,s$, as A .

We partition $Y_{i,k}$, given by (4.7), into

$$Y_{i,k} = \begin{pmatrix} Y_{1,i,k} \\ \vdots \\ Y_{s,i,k} \end{pmatrix}$$

for $i=0,1,\dots,p_k$, so that the m_j by N_k matrices $Y_{j,i,k}$ are given by

$$Y_{j,i,k} = (y_{j,k-i} : y_{j,k-i+d} : \dots : y_{j,k-i+(N_k-1)d})$$

for $j=1, \dots, s$. We define the $m_j p_k$ by N_k matrix $H_{j,k}$ by

$$H_{j,k} = \begin{pmatrix} \tilde{Y}_{j,1,k} \\ \vdots \\ Y_{j,p_k,k} \end{pmatrix}$$

for $j=1, \dots, s$ and the mp_k by N_k matrix H_k by

$$H_k = \begin{pmatrix} H_{1,k} \\ \vdots \\ H_{s,k} \end{pmatrix}.$$

Let P' be the permutation matrix such that

$$H_k = P' Y'_k.$$

We also define the m_i by $m_j p_k$ matrices $\Xi_{ij}(k)$, $i, j=1, \dots, s$, by the relation

$$\Phi(k) = \begin{pmatrix} \Xi_{11}(k) & \dots & \Xi_{1s}(k) \\ \vdots & & \vdots \\ \Xi_{s1}(k) & \dots & \Xi_{ss}(k) \end{pmatrix} P'$$

where as before $\Phi(k) = (\phi_1(k) : \dots : \phi_{p_k}(k))$. It follows that $\Xi_{ij}(k) = (\phi_{1,ij}(k) : \phi_{2,ij}(k) : \dots : \phi_{p_k,ij}(k))$ and hence in the simple correlated case we have that $\Xi_{ij}(k) = 0$ if $i \neq j$.

Least Gramian estimates

THEOREM 5.4. If $M(Y_{j,0,k}^T) \subset M(H_{j,k}^T)$, $j=1, \dots, s$ and F is the range space of $\bigoplus_{j=1}^s (Y_{j,0,k} H_k^T (H_{j,k} H_k^T)^{-})$ then there exist an LG estimator $\hat{\Phi}_S^{LG}(k)$ of $\Phi(k) \in F$ in the PSMA model. An estimator can taken to be

$$(5.5) \quad \hat{\Phi}_S^{LG}(k) = \left(\bigoplus_{j=1}^s Y_{j,0,k} H_k^T (H_{j,k} H_k^T)^{-} \right) P'.$$

PROOF. Write $\Phi(k) = \hat{\Phi}_S^{LG} - \Delta$ where $\hat{\Phi}_S^{LG}$ is given by (5.5) and define $\hat{\Xi}_{jj} = Y_{j,0,k} H_k^T (H_{j,k} H_k^T)^{-}$. Then

$$\hat{\phi}_S^{LG}(k)Y'_k = \left(\bigoplus_{j=1}^S \hat{\Xi}_{jj} \right) H_k = \begin{pmatrix} \hat{\Xi}_{11}(k)H_{1,k} \\ \vdots \\ \hat{\Xi}_{ss}(k)H_{s,k} \end{pmatrix}$$

and hence

$$(5.6) \quad (Y_k - \hat{\phi}_S^{LG}(k)Y'_k)(\Delta Y'_k)^T = (Y_k - \begin{pmatrix} \hat{\Xi}_{11}(k)H_{1,k} \\ \vdots \\ \hat{\Xi}_{ss}(k)H_{s,k} \end{pmatrix}) H_k^T P' \Delta^T.$$

By assumption, $Y_{j,0,k} = AH_{j,k}$ for some matrix A and thus

$$\hat{\Xi}_{jj}(k)H_{j,k}H_k^T = Y_{j,0,k}H_k^T(H_{j,k}H_k^T)^{-1}H_{j,k}H_k^T = Y_{j,0,k}H_k^T.$$

Hence the expression (5.6) is equal to zero independent of Δ and we infer that

$$\begin{aligned} (Y_k - \hat{\phi}(k)Y'_k)(Y_k - \hat{\phi}(k)Y'_k)^T &= (Y_k - \hat{\phi}_S^{LG}(k)Y'_k)(Y_k - \hat{\phi}_S^{LG}(k)Y'_k)^T \\ &\quad + \Delta Y'_k(\Delta Y'_k)^T \\ &\geq (Y_k - \hat{\phi}_S^{LG}(k)Y'_k)(Y_k - \hat{\phi}_S^{LG}(k)Y'_k)^T \end{aligned}$$

which proves the theorem. □

An estimate $\hat{\Sigma}_S^{LG}(k)$ of $\Sigma(k)$ is given by

$$\begin{aligned} (5.7) \quad \hat{\Sigma}_S^{LG}(k) &= (Y_k - \begin{pmatrix} \hat{\Xi}_{11}(k)H_{1,k} \\ \vdots \\ \hat{\Xi}_{ss}(k)H_{s,k} \end{pmatrix}) (Y_k - \begin{pmatrix} \hat{\Xi}_{11}(k)H_{1,k} \\ \vdots \\ \hat{\Xi}_{ss}(k)H_{s,k} \end{pmatrix})^T / f_k^S \\ &= \begin{pmatrix} \hat{\Sigma}_{11} & \dots & \hat{\Sigma}_{1s} \\ \vdots & & \vdots \\ \hat{\Sigma}_{s1} & \dots & \hat{\Sigma}_{ss} \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} \hat{\Sigma}_{ij}(k) &= Y_{i,0,k}(I_{N_k} - H_k^T(H_{i,k}H_k^T)^{-1}H_{i,k})Y_{j,0,k}^T / f_k^S \\ &= Y_{i,0,k}(I_{N_k} - H_{j,k}^T(H_kH_{j,k}^T)^{-1}H_{j,k})Y_{j,0,k}^T / f_k^S. \end{aligned}$$

Here $f_k^S = N_k - r_k^S$ and $r_k^S = \text{tr}(H_k^T (H_{1,k} H_k^T)^{-1} H_{1,k})$.

Yule-Walker estimates

Define

$$\Gamma_{ij}^{ab}(k) = E(y_{a,k-i} y_{b,k-j}^T)$$

where $y_{a,k}$ is defined from the relation $y_k = \text{vec}(y_{1,k} : \dots : y_{s,k})$. The matrix $\Gamma_{ij}^{ab}(k)$ is the (a,b) 'th block of order (m_a, m_b) of $\Gamma_{ij}(k)$, that is

$$\Gamma_{ij}(k) = \begin{pmatrix} \Gamma_{ij}^{11}(k) & \dots & \Gamma_{ij}^{1s}(k) \\ \vdots & & \vdots \\ \Gamma_{ij}^{s1}(k) & \dots & \Gamma_{ij}^{ss}(k) \end{pmatrix}.$$

In the simple correlated case the following result connects the values of the covariance at different times.

THEOREM 5.5. The second order moments $\Gamma_{ij}^{ab}(k)$ of the PMSAR(Q) process $\{y_k = \text{vec}(y_{1,k} : \dots : y_{s,k}) \in \mathbb{R}^m, k \in \mathbb{Z}\}$ generated by (2.2) satisfy the following relations.

- (i) $\Gamma_{i0}^{ab}(k) = \sum_{j=1}^{q_{b,k}} \Gamma_{ij}^{ab}(k) \phi_{j,bb}(k)^T, \quad i > 0, a, b = 1, \dots, s$
- (ii) $\Gamma_{00}^{ab}(k) = \sum_{j=1}^{q_{b,k}} \Gamma_{0j}^{ab}(k) \phi_{j,bb}(k)^T + \Sigma_{ab}(k), \quad a, b = 1, \dots, s.$

PROOF. We have using (2.2)

$$y_{a,k-i} e_{b,k}^T = y_{a,k-i} (y_{b,k} - \sum_{j=1}^{q_{b,k}} \phi_{j,bb}(k) y_{b,k-j})^T$$

and by taking expectations we have

$$E(y_{a,k-i} e_{b,k}^T) = \Gamma_{i0}^{ab}(k) - \sum_{j=1}^{q_{b,k}} \Gamma_{ij}^{ab}(k) \phi_{j,bb}(k)^T$$

the left hand side of which is zero for $i > 0$ and $\Sigma_{ab}(k)$ for $i=0$. $\square$

The results of the theorem extend to the multivariate and periodic case, results given by Risager (1981, Theorem 3.2) for the simple correlated

univariate autoregressive process.

5.2.2. Estimation in the PBMAR process

Another useful partitioning of $\Phi(k)$ is in terms of the m_j by mp_k matrices $\Omega_j(k)$, $j=1, \dots, s$, given by the relation

$$\Phi(k) = \begin{pmatrix} \Omega_1(k) \\ \vdots \\ \Omega_s(k) \end{pmatrix}.$$

When the covariance matrices are block diagonal we partition $\Sigma(k)$ into

$$\Sigma(k) = \begin{pmatrix} \Sigma_{11}(k) & \dots & \Sigma_{1s}(k) \\ \vdots & & \vdots \\ \Sigma_{s1}(k) & \dots & \Sigma_{ss}(k) \end{pmatrix}.$$

so that $\Sigma_{uv} = 0$ for $u \neq v$.

Least Gramian estimates

We minimize

$$\begin{aligned} & B_{\frac{m}{s}} ((Y_k - \Phi(k)Y'_k)(Y_k - \Phi(k)Y'_k)^T) \\ & \equiv \bigoplus_{i=1}^s (Y_{i,0,k} - \Omega_i(k)Y'_k)(Y_{i,0,k} - \Omega_i(k)Y'_k)^T. \end{aligned}$$

Since the i 'th diagonal block involves just Ω_i , the minimization can be done block-wise giving, analogously to (4.3),

$$\hat{\Omega}_i(k) = Y_{i,0,k} Y_k'^T (Y_k' Y_k')^{-}$$

and thus

$$\hat{\Phi}_B(k) \equiv \begin{pmatrix} \hat{\Omega}_1(k) \\ \vdots \\ \hat{\Omega}_s(k) \end{pmatrix} = Y_k Y_k'^T (Y_k' Y_k')^{-}.$$

An estimate of $\Sigma_{ii}(k)$ is given by

$$\hat{\Sigma}_{ii}(k) = Y_{i,0,k} (I_{N_k} - Y_k^T (Y_k' Y_k')^{-1} Y_k) Y_{i,0,k}^T / (N_k - r_k^0)$$

where $r_k^0 = \text{tr}(Y_k^T (Y_k' Y_k')^{-1} Y_k)$, so that

$$(5.8) \quad \hat{\Sigma}_B^{LG}(k) = B_{\underline{m}_S} (Y_k (I_{N_k} - Y_k^T (Y_k' Y_k')^{-1} Y_k) Y_k^T / (N_k - r_k^0)).$$

5.2.3. Estimation in the PSBMAR process

Let, in the simple correlated model, the covariance matrices $\Sigma(k)$, $k=1, \dots, d$, be block diagonal with block sizes identical to the dimensions of the innovation correlated subprocesses.

Least Gramian estimates

In this case we minimize

$$\begin{aligned} & B_{\underline{m}_S} \left((Y_k - \begin{pmatrix} \Xi_{11}(k) H_{1,k} \\ \vdots \\ \Xi_{ss}(k) H_{s,k} \end{pmatrix}) (Y_k - \begin{pmatrix} \Xi_{11}(k) H_{1,k} \\ \vdots \\ \Xi_{ss}(k) H_{s,k} \end{pmatrix})^T \right) \\ &= \bigoplus_{i=1}^s (Y_{i,0,k} - \Xi_{ii}(k) H_{i,k}) (Y_{i,0,k} - \Xi_{ii}(k) H_{i,k})^T \end{aligned}$$

for $k=1, \dots, d$.

Since the i 'th factor in the direct sum involves just $\Xi_{ii}(k)$, the minimization can be done factorwise and thus, analogously to (4.3), we obtain the LG-estimators

$$\hat{\Xi}_{ii}(k) = Y_{i,0,k} H_{i,k}^T (H_{i,k} H_{i,k}^T)^{-1},$$

and

$$\hat{\Sigma}_{ii}(k) = Y_{i,0,k} (I_{N_k} - H_{i,k}^T (H_{i,k} H_{i,k}^T)^{-1} H_{i,k}) Y_{i,0,k}^T / (N_k - r_k^{SB})$$

for $i=1, \dots, s$, where $r_k^{SB} = \text{tr}(H_{i,k}^T (H_{i,k} H_{i,k}^T)^{-1} H_{i,k})$ and thus

$$(5.9) \quad \hat{\Sigma}_{SB}(k) = B_{\underline{m}_S} (Y_k (\bigoplus_{i=1}^s (I_{N_k} - H_{i,k}^T (H_{i,k} H_{i,k}^T)^{-1} H_{i,k})) Y_k^T / (N_k - r_k^{SB})).$$

6. DIAGNOSTIC CHECKING

In this section we consider statistics suitable for verifying or dismissing the PMAR process as a suitable model for observed time series. Before we consider in section 6.2 the most common situation where the parameters are not known, we look at a somewhat hypothetical situation where the regression parameters are supposed to be known.

6.1. KNOWN REGRESSION PARAMETERS

When the regression parameters $\Phi(k) = (\phi_1(k) : \dots : \phi_{n_k}(k))$, $k=1, \dots, d$, are known in the model given by (2.1), the verification procedure or diagnostic checking concerns the assumptions on the innovation process $\{e_k\}$. For example, one may be interested in checking the assumption of normality in the PGMAR process, block-diagonal covariance matrices in the PBMAR process, time homogeneity in the P_IMAR process or checking the known regression parameters by regarding the goodness of fit to the given values.

Although the results given in this section are valid for any sequence $\{e_k\}$ with specified properties, we shall think of the process as obtained from the observable process $\{y_k\}$ by the relation $e_k = y_k - \Phi(k)y'_k$.

We define for $k = 1, \dots, d$,

$$s_{ij}^{ab}(k) = N_k^{-1} \sum_{\ell=1}^{N_k} e_{a,k-i+(\ell-1)d} e_{b,k-j+(\ell-1)d}^T$$

where $e_{a,k}$ is defined in analogy with $y_{a,k}$, for $a, b = 1, \dots, s$. For brevity in notation we define

$$s_{ij}^{ab}(k) = \text{vec}(S_{ij}^{ab}(k)) = N_k^{-1} \sum_{\ell=1}^{N_k} e_{a,k-i+(\ell-1)d} \otimes e_{b,k-j+(\ell-1)d}^T$$

Here $S_{ij}^{ab}(k)$ is of order (m_a, m_b) and is the (a, b) 'th block of

$$S_{ij}(k) = \begin{pmatrix} S_{ij}^{11}(k) & \dots & S_{ij}^{1s}(k) \\ \vdots & & \vdots \\ S_{ij}^{s1}(k) & \dots & S_{ij}^{ss}(k) \end{pmatrix} = N_k^{-1} \sum_{\ell=1}^{N_k} e_{k-i+(\ell-1)d} e_{k-j+(\ell-1)d}^T$$

which is a natural extension to periodic processes of the autocovariance function of lag $i-j$ for aperiodic ($d=1$) processes. We shall construct a number of portmanteau lack of fit statistics based on $S_{ij}(k)$, or $s_{ij}(k) \equiv \text{vec}(S_{ij}(k))$, $i, j = 1, \dots, h$. Some of these involve the autocovariance $s_{ij}^{aa}(k)$ for the a 'th series which shall be used for measuring the consistency of data with the model for the a 'th series. Others involve the cross covariance $s_{ij}^{ab}(k)$ between the a 'th and b 'th series, which shall be used for measuring the consistency of data with the model involving both the a 'th and the b 'th series.

The first results concern the two first central moments of these statistics and the limiting distribution when the length of the realization increases to infinity. For an arbitrary random matrix, we let $E(X)$ denote componentwise expectation, and $\mathcal{D}(X)$ denotes the arrangement $E\{(X - E(X)) \square (X - E(X))\}$ of the second order central moments.

THEOREM 6.1. Let $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ be a PMAR(p_d^T) process given by (2.1) with known regression parameters $\{\phi_i(k), i=1, \dots, p_k, k=1, \dots, d\}$. Suppose $\{e_k\}$ is an independent sequence where $e_k = y_k - \phi(k)y'_k$; then the following results hold:

$$(i) \quad E(S_{ij}(k)) = \begin{cases} 0 & i \neq j \\ \Sigma(k-i) & i=j. \end{cases}$$

$$(ii) \quad \mathcal{D}(S_{ij}(k)) = \begin{cases} N_k^{-1}(\Sigma(k-i) \otimes \Sigma(k-j)) & i \neq j \\ N_k^{-1}(E(e_{k-i}^{<2>} e_{k-i}^{<2>T}) - \Sigma(k-i) \square \Sigma(k-i)) & i=j. \end{cases}$$

(iii) If $i \leq j < L$, then $S_{ij}(k)$ and $S_{iL}(k)$ are uncorrelated.

(iv) If $i \neq j$, then

$$N_k^{1/2} s_{ij}(k) \xrightarrow{L} N_m^2(0, \Sigma(k-i) \otimes \Sigma(k-j)),$$

as $n \rightarrow \infty$.

(v) If in addition $E((e_{k-i}^T e_{k-i})^2) < +\infty$, then

$$N_k^{-1/2} (s_{ii}(k) - \text{vec}(\Sigma(k-i))) \xrightarrow{L} N_m^2(0, E(e_{k-i}^{<2>} e_{k-i}^{<2>T}) - \Sigma(k-i) \square \Sigma(k-i)),$$

as $n \rightarrow \infty$.

PROOF. See Appendix E.

The result in Theorem 6.1 (iii) can be written

$$N_k^{-1/2} S_{ij}(k) \xrightarrow{L} N_{m,m}(0, \Sigma(k-i), \Sigma(k-j)),$$

or

$$N_k^{-1/2} \Sigma(k-i)^{-1/2} S_{ij}(k) \Sigma(k-j)^{-1/2} \xrightarrow{L} N_{m,m}(0, I_m, I_m),$$

as $N_k \rightarrow \infty$.

It follows from Theorem 6.1 (i) and (ii) that $S_{ii}(k)$ is a consistent estimate of $\Sigma(k-i)$. Hence

$$N_k^{-1/2} S_{ii}(k)^{-1/2} S_{ij}(k) S_{jj}(k)^{-1/2} \xrightarrow{L} N_{m,m}(0, I_m, I_m)$$

as $N_k \rightarrow \infty$.

As an immediate result of Theorem 6.1, we have the limiting normal distribution of the auto- and crosscovariance statistics of the subprocesses.

COROLLARY 6.1. Under the same hypothesis as in Theorem 6.1 the following results hold:

$$(i) \quad E(S_{ij}^{ab}(k)) = \begin{cases} 0 & i \neq j \\ \Sigma_{ab}(k-i) & i=j, \end{cases}$$

$$(ii) \quad \mathcal{D}(S_{ij}^{ab}(k)) = \begin{cases} N_k^{-1} \Sigma_{aa}(k-i) \otimes \Sigma_{bb}(k-j) & i \neq j \\ N_k^{-1} (E(e_{a,k-i} e_{a,k-i}^T \otimes e_{b,k-i} e_{b,k-i}^T) - \Sigma_{ab}(k-i) \square \Sigma_{ab}(k-i)) & i=j, \end{cases}$$

(iii) If the a'th and b'th innovation subprocesses $\{e_{a,k}\}$ and $\{e_{b,k}\}$ are uncorrelated, then for arbitrary integers i and j it holds that

$$\mathcal{D}(S_{ij}^{ab}(k)) = N_k^{-1} \Sigma_{aa}(k-i) \otimes \Sigma_{bb}(k-j)$$

(iv) For $i \neq j$ it holds that

$$N_k^{\frac{1}{2}} s_{ij}^{ab}(k) \xrightarrow{L} N_{m_a m_b} (0, \Sigma_{aa}(k-i) \otimes \Sigma_{bb}(k-j))$$

as $n \rightarrow \infty$.

(v) If in addition $E((e_{k-i}^T e_{k-i})^2) < +\infty$,

$$N_k^{\frac{1}{2}} (s_{ii}^{ab}(k) - \text{vec}(\Sigma_{ab}(k-i))) \xrightarrow{L} N_{m_a m_b} (0, V), \quad n \rightarrow \infty,$$

where $V = E(e_{a,k-i}^T e_{a,k-i} \otimes e_{b,k-i}^T e_{b,k-i} - \Sigma_{ab}(k-i) \square \Sigma_{ab}(k-i))$.

(vi) If the a 'th and b 'th innovation subprocesses are uncorrelated, we have

$$N_k^{\frac{1}{2}} s_{ii}^{ab}(k) \xrightarrow{L} N_{m_a m_b} (0, \Sigma_{a,a}(k-i) \otimes \Sigma_{b,b}(k-i)), \quad n \rightarrow \infty.$$

PROOF. (i): Let $T_{m,n}^{(1,\rho)}$ be such that $T_{m,n}^{(1,\rho)} (B \otimes A) T_{m,n}^{(\rho,1)} = A \otimes B$ where $T_{m,n}^{(\rho,1)} = (T_{m,n}^{(1,\rho)})^T$ and A is m -square and B is n -square. It can be shown that

$$(6.1) \quad s_{ij}(k) = \left(\bigoplus_{i=1}^s T_{m,m_i}^{(1,\rho)} \right) \left(\bigoplus_{i=1}^s \left(\bigoplus_{j=1}^s T_{m_i,m_j}^{(1,\rho)} \right) \right) \\ \text{vec}(s_{ij}^{11}(k) : \dots : s_{ij}^{s1}(k) : \dots : s_{ij}^{1s}(k) : \dots : s_{ij}^{ss}(k))$$

and

$$(6.4) \quad \text{vec}(\Sigma(k)) = \left(\bigoplus_{i=1}^s T_{m,m}^{(1,\rho)} \right) \left(\bigoplus_{i=1}^s \left(\bigoplus_{j=1}^s T_{m_i,m_j}^{(1,\rho)} \right) \right) \\ \text{vec}(\text{vec} \Sigma_{11}(k) : \dots : \text{vec} \Sigma_{s1}(k) : \dots : \text{vec} \Sigma_{1s}(k) : \dots : \text{vec} \Sigma_{ss}(k)).$$

Hence it follows from Theorem 6.1 (i) that $E(s_{ij}^{ab}(k)) = 0$ if $i \neq j$ and

$E(s_{ij}^{ab}(k)) = \text{vec}(\Sigma_{ab}(k-i))$ if $i=j$. This proves (i).

(ii): It follows from (6.1) that $\mathcal{D}(s_{ij}^{ab}(k))$ is the $(a+(b-1)s)$ 'th diagonal block - of size $m_a m_b$ -square - of

$$\left(\bigoplus_{i=1}^s \left(\bigoplus_{j=1}^s T_{m_j,m_i}^{(1,\rho)} \right) \right) \left(\bigoplus_{i=1}^s T_{m_i,m}^{(1,\rho)} \right) \mathcal{D}(s_{ij}(k)) \left(\bigoplus_{i=1}^s T_{m_i,m}^{(\rho,1)} \right) \left(\bigoplus_{i=1}^s \left(\bigoplus_{j=1}^s T_{m_i,m_j}^{(\rho,1)} \right) \right)$$

Writing

$$N_k \mathcal{D}(s_{ij}(k)) \\ = \left(\bigoplus_{u=1}^s T_{m,m_u}^{(1,\rho)} \right) \left(\bigoplus_{u=1}^s \left(\bigoplus_{v=1}^s T_{m_u,m_v}^{(1,\rho)} \right) \right) K \left(\bigoplus_{u=1}^s \left(\bigoplus_{v=1}^s T_{m_u,m_v}^{(\rho,1)} \right) \right) \left(\bigoplus_{u=1}^s T_{m,m_u}^{(\rho,1)} \right)$$

and partitioning K into

$$K = \begin{pmatrix} K_{11} & \cdots & K_{1s} \\ \vdots & & \vdots \\ K_{s1} & \cdots & K_{ss} \end{pmatrix}$$

where K_{uv} is of order mm_u by mm_v , it can be shown by means of Theorem 6.1 (ii) that

$$K_{uv} = \begin{pmatrix} \Sigma_{11}(k-i) \otimes \Sigma_{uv}(k-j) & \cdots & \Sigma_{1s}(k-i) \otimes \Sigma_{uv}(k-j) \\ \vdots & & \vdots \\ \Sigma_{s1}(k-i) \otimes \Sigma_{uv}(k-j) & \cdots & \Sigma_{ss}(k-i) \otimes \Sigma_{uv}(k-j) \end{pmatrix}$$

for $i \neq j$ and

$$K_{uv} = E \begin{pmatrix} K'_{11} \otimes K'_{uv} & \cdots & K'_{1s} \otimes K'_{uv} \\ \vdots & & \vdots \\ K'_{s1} \otimes K'_{uv} & \cdots & K'_{ss} \otimes K'_{uv} \end{pmatrix} - \begin{pmatrix} \Sigma_{lu}(k-i) \square \Sigma_{lv}(k-i) & \cdots & \Sigma_{lu}(k-i) \square \Sigma_{sv}(k-i) \\ \vdots & & \vdots \\ \Sigma_{su}(k-i) \square \Sigma_{lv}(k-i) & \cdots & \Sigma_{su}(k-i) \square \Sigma_{sv}(k-i) \end{pmatrix},$$

for $i=j$, where $K'_{pq} = e_{p,k-i} e_{q,k-i}^T$. From these expressions we infer that the $(a+(b-1)s)$ 'th diagonal block of K is

$$\Sigma_{aa}(k-i) \otimes \Sigma_{bb}(k-j).$$

if $i \neq j$ and

$$E(K'_{aa} \otimes K'_{bb}) - \Sigma_{ab}(k-i) \square \Sigma_{ab}(k-i)$$

if $i=j$. From Theorem 6.1 (ii) the result then follows.

The results in (iii) and (iv) follows from (i) and (ii) and Theorem 6.1 (iii) and (iv). This completes the proof.

Since the asymptotic distributions of $S_{ij}(k)$ and $S_{ij}^{ab}(k)$, given by Theorem 6.1 and Corollary 6.1, involve second and forth order moments of the innovation process, these statistics can be used for verification purposes only if these moments are known. When the moments are unknown, it is customary to combine the covariance matrix with estimates of the moments so that the resulting statistic is independent (at least asymptotically) of the moments

mentioned. Portmanteau statistics are therefore defined in terms of correlations instead of covariances.

As seen in section 3, the periodic MAR process can be regarded as a non-periodic MAR process by considering a whole period of the process.

Define, with $N = \min(N_1, \dots, N_d) = [n/d]$,

$$V_{ij} = N^{-1} \sum_{\ell=1}^N v_{1-i+\ell} v_{1-j+\ell}^T$$

where as before $v_k = \text{vec}(e_{1+(k-1)d} : \dots : e_{kd})$.

When, in the deperiodized process, the regression parameters are known, the innovations $\{u_k\}$ can be calculated from the observable variables $\{x_k\}$. It is then natural to construct a statistic based on these instead of the process $\{v_k\}$. We define

$$U_{ij} = N^{-1} \sum_{\ell=1}^N u_{1-i+\ell} u_{1-j+\ell}^T$$

with the following relation to V_{ij} :

$$U_{ij} = \Lambda V_{ij} \Lambda^T.$$

THEOREM 6.2. Let $\{y_k \in \mathbb{R}^m, k \in \mathbb{Z}\}$ be a PMAR(p_d^T) process given by (2.1) with known regression parameters $\{\phi_i(k), i=1, \dots, p_k, k=1, \dots, d\}$. Let θ be as in the proof of Theorem 3.1. Suppose $\{e_k\}$ is an independent sequence and define $u_k = \Lambda v_k = \Lambda \text{vec}(e_{1+(k-1)d} : \dots : e_{kd})$ and $e_k = y_k - \phi(k)y_k'$; then

$$(i \ a) \quad E(V_{ij}) = \begin{cases} 0 & \text{if } i \neq j \\ \bigoplus_{k=1}^d \Sigma(k) & \text{if } i=j. \end{cases}$$

$$(i \ b) \quad E(U_{ij}) = \begin{cases} 0 & \text{if } i \neq j \\ \theta & \text{if } i=j. \end{cases}$$

$$(ii \ a) \quad \mathcal{D}(V_{ij}) = \begin{cases} N^{-1} \left(\bigoplus_{k=1}^d \Sigma(k) \right) \otimes \left(\bigoplus_{k=1}^d \Sigma(k) \right) & \text{if } i \neq j \\ N^{-1} \left(E(v_{k-i}^{<2>} v_{k-i}^{<2>T}) - \left(\bigoplus_{k=1}^d \Sigma(k) \right) \square \left(\bigoplus_{k=1}^d \Sigma(k) \right) \right) & \text{if } i=j. \end{cases}$$

$$(ii \ b) \quad \mathcal{D}(U_{ij}) = \begin{cases} N^{-1} (\theta \otimes \theta) & \text{if } i \neq j \\ N^{-1} (E(u_{k-i}^{<2>} u_{k-i}^{<2>T}) - (\theta \square \theta)) & \text{if } i=j \end{cases}$$

(iii a) If $i \leq j < \ell$, then V_{ij} and $V_{i\ell}$ are uncorrelated.

(iii b) If $i \leq j < \ell$, then U_{ij} and $U_{i\ell}$ are uncorrelated.

(iv a) If $i \neq j$, then

$$N^{\frac{1}{2}} V_{ij} \xrightarrow{L} N_{md,md} \left(0, \bigoplus_{k=1}^d \Sigma(k-i), \bigoplus_{k=1}^d \Sigma(k-j) \right),$$

as $n \rightarrow \infty$.

(iv b) If $i \neq j$, then

$$N^{\frac{1}{2}} U_{ij} \xrightarrow{L} N_{md,md} (0, \theta, \theta),$$

as $n \rightarrow \infty$.

(v a) If in addition $\{v_k\}$ is fourth order stationary with finite fourth order moment $E((\sum_{k=1}^d e_k^T e_k)^2) < +\infty$, then

$$N^{\frac{1}{2}} \text{vec}(V_{ii} - (\bigoplus_{k=1}^d \Sigma(k))) \xrightarrow{L} N_{m^2 d^2} (0, E(v_k^{<2>} v_k^{<2>T}) - (\bigoplus_{k=1}^d \Sigma(k)) \square (\bigoplus_{k=1}^d \Sigma(k))),$$

as $n \rightarrow \infty$.

(v b) If, in addition, $\{v_k\}$ is fourth order stationary with finite fourth order moment $E((\sum_{k=1}^d e_k^T e_k)^2) < +\infty$, then

$$N^{\frac{1}{2}} \text{vec}(U_{ii} - \theta) \xrightarrow{L} N_{m^2 d^2} (0, E(u_k^{<2>} u_k^{<2>T}) - \theta \square \theta)$$

as $n \rightarrow \infty$.

PROOF. See Appendix E. □

6.2. UNKNOWN REGRESSION PARAMETERS

When the regression parameters are unknown, a portmanteau lack of fit statistic similar to the one considered in last section can be constructed based on the empirical residuals obtained with estimates of the parameters inserted for the true values. The distribution of this portmanteau lack of

fit statistic depends on the distribution of the original process as well as of the method used for obtaining estimates of the parameters.

For a large class of estimators of the parameters, the asymptotic distribution of the portmanteau lack of fit statistic is independent of the estimation procedure and the distribution of the process.

We now give the large-sample distribution of the portmanteau lack of fit statistic for unknown regression parameters when these are estimated with least Gramian estimates.

THEOREM 6.3. For $h > p$ it holds that

$$N \sum_{\ell=2}^{h+1} \hat{U}_{11}^{-1/2} \hat{U}_{1\ell} \hat{U}_{11}^{-1} \hat{U}_{\ell 1} \hat{U}_{11}^{-1/2} \in W_{md}(md(h-p), I_{md})$$

asymptotically, as $N \rightarrow \infty$, where $\hat{U}_{1\ell} = N^{-1} \sum_{u=1}^N (x_u - \hat{\Psi}_{x'_u})(x_{1-\ell+u} - \hat{\Psi}_{x'_{1-\ell+u}})^T$.

PROOF. See Appendix F. □

For the original PMAR process a similar result holds.

THEOREM 6.4. For $h > p_k$ it holds that asymptotically, as $n \rightarrow \infty$, that

$$N_k \sum_{\ell=2}^{h+1} \hat{S}_{11}(k)^{-1/2} \hat{S}_{1\ell}(k) \hat{S}_{11}(k)^{-1} \hat{S}_{\ell 1}(k) \hat{S}_{11}(k)^{-1/2} \in W_m(m(h-p_k), I_m)$$

where $\hat{S}_{ij}(k) =$

$$N_k^{-1} \sum_{u=1}^{N_k} (y_{k-i+(u-1)d}^{-\hat{\Phi}(k-i)} y'_{k-i+(u-1)d}) (y_{k-j+(u-1)d}^{-\hat{\Phi}(k-j)} y'_{k-j+(u-1)d})^T$$

PROOF. See Appendix F. □

The result of Theorem 6.3 can be used for testing consistency of data with the deperiodized process and the statistic in Theorem 6.4 is used for the corresponding test in the original process. Since the statistic is matrix valued, it might be difficult to use as a check of consistency with the model since a critical region is not naturally given. A way to use the result is to construct univariate statistics such as

$$N \sum_{\ell=2}^{h+1} \text{tr}(\hat{U}_{11}^{-1/2} \hat{U}_{1\ell} \hat{U}_{11}^{-1} \hat{U}_{\ell 1} \hat{U}_{11}^{-1/2}) \in \chi^2((md)^2(h-p))$$

and

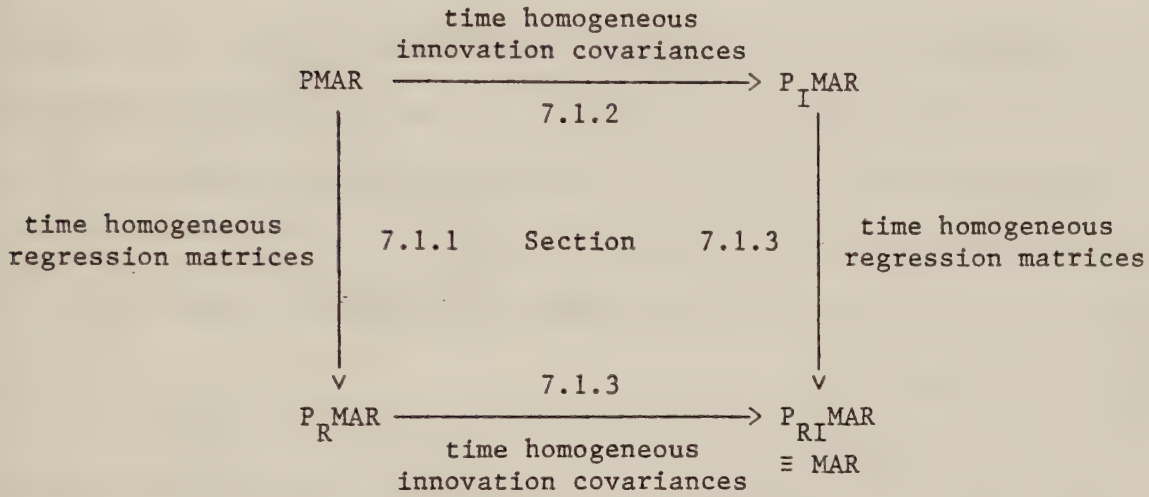
$$|N \sum_{\ell=2}^{h+1} \hat{U}_{11}^{-1/2} \hat{U}_{1\ell} \hat{U}_{11}^{-1} \hat{U}_{\ell 1} \hat{U}_{11}^{-1/2}| \in U(md(h-p), md)$$

where $U(p, m)$ is the distribution of the determinant of a $W_m(p, I_m)$ distributed random matrix.

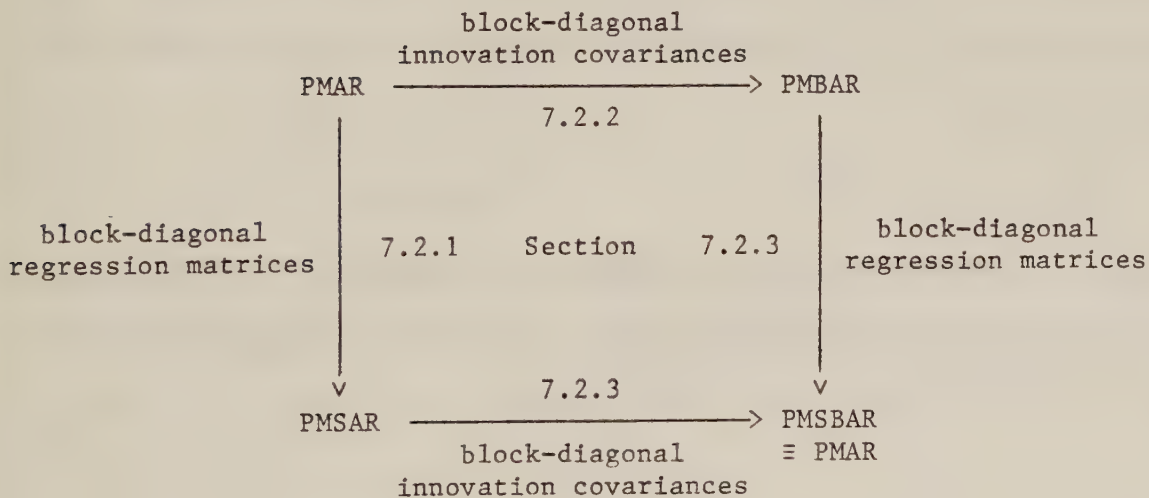
7. HYPOTHESIS TESTING

We shall deal with the two following schemes for reducing the general PMAR process.

Time homogeneity



Reduced correlation structure



The main part of the test statistics proposed in this section consists of analogies of test statistics derived as likelihood ratio statistics for the analogous multivariate linear normal model (MLNM).

7.1. TIME HOMOGENEITY

In this section we consider test statistics for testing hypothesis concerning homogeneity in time of the regression matrices and the innovation covariances.

7.1.1. Homogeneous regression in the PMAR process

A problem in the MLNM case which is similar to the problem of testing equality of the regression matrices, for $k=1, \dots, d$, is the testing of equality of several means in a set of normally distributed samples.

The test statistic used for testing homogeneous regression versus the more general PMAR model is

$$(7.1) \quad \delta_R = \frac{\prod_{k=1}^d |\hat{\Sigma}^{LG}(k)|^{\frac{1}{2}(N_k - r_k)}}{\prod_{k=1}^d |\hat{\Sigma}_R^{LG}(k)|^{\frac{1}{2}(N_k - r_k)}}$$

where $\hat{\Sigma}^{LG}(k)$ and $\hat{\Sigma}_R^{LG}(k)$ are the least Gramian estimates in the general model and the model with time-independent regression matrices, respectively.

The hypothesis

$$H_R: \Phi(1) = \dots = \Phi(d)$$

is rejected for small values of δ_R . Under the assumption that the regression matrices are independent of time, the asymptotic distribution of $-2\ln\delta_R$ is $\chi^2((\bar{p}d - p')m^2)$ where $\bar{p} = \sum_{k=1}^d p_k/d$.

7.1.2. Homogeneous innovations in the PMAR process

The case when the error process has a homogeneous covariance operator (i.e. $\Sigma(k) = \Sigma$, $k=1, \dots, d$) is of special interest since one may in some situations assume that the error process represents the environmental im-

pact on the process and that such an impact does not vary with time.

To find a suitable test statistic we look at a similar situation in the MLNM case, namely criteria for testing equality of several covariance matrices in a set of normally distributed samples.

To test the hypothesis

$$H_I: \Sigma(1) = \dots = \Sigma(d)$$

of equality of the innovation covariances we reject this hypothesis for small values of

$$(7.2) \quad \delta_I = \frac{\prod_{k=1}^d |\hat{\Sigma}^{LG}(k)|^{\frac{1}{2}(N_k - r_k)}}{|\hat{\Sigma}_I^{LG}|^{\frac{1}{2}(n-r)}}$$

where $n = \sum_{k=1}^d N_k$ and $r = \sum_{k=1}^d r_k$ and where $\hat{\Sigma}^{LG}(k)$ and $\hat{\Sigma}_I^{LG}$ are the LG estimates under the two models.

The test statistic for testing equality of several covariance matrices in the MLNM case was given by Anderson (1984) (with r_k denoting the row order of Z_k). When r_k is the (row) rank of Z_k , the test which rejects H_I for small values of

$$\delta_I^* = \frac{\prod_{k=1}^d |\hat{\Sigma}^{LG}(k)^*|^{\frac{1}{2}(N_k - r_k)}}{|\hat{\Sigma}_I^{LG*}|^{\frac{1}{2}(n-r)}}$$

is unbiased for all alternatives (Holmquist (1985c)).

Under the assumption that the covariance matrices are equal, the asymptotic distribution of $-2\ln\delta_I$ is $\chi^2_{\frac{1}{2}((d-1)m(m+1))}$ as $n \rightarrow \infty$.

7.1.3. Time homogeneity in the PMAR process

If the regression matrices as well as the innovation covariances in the PMAR process are independent of time, the periodic MAR reduces to the ordinary MAR process.

Homogeneous innovation covariances in the P_R MAR process

A test statistic which can be used for testing homogeneous innovations in the homogeneous regression PMAR model is

$$(7.3) \quad \delta_{I|R} = \frac{\prod_{k=1}^d |\hat{\Sigma}_R^{LG}(k)|^{\frac{1}{2}(N_k - r_k)}}{|\hat{\Sigma}_{RI}^{LG}|^{\frac{1}{2}(n-r)}}$$

where $\hat{\Sigma}_R^{LG}(k)$ is the LG estimate when the regression matrices are independent of time and $\hat{\Sigma}_{RI}^{LG}$ is given by (5.4). The hypothesis H_I is rejected for small values of $\delta_{I|R}$.

Under the assumption of time homogeneous innovation covariances, the asymptotic distribution of $-2\ln\delta_{I|R}$ is $\chi^2(\frac{1}{2}(d-1)m(m+1))$ as n tends to infinity.

The statistic is an adaption to the present case of the likelihood ratio statistic in the MLNM case for testing equality of covariances given equality of means in several normal populations.

Homogeneous regression matrices in the P_I MAR process

This situation corresponds to testing of equality of mean vectors in the MLNM case when we have assumed that the covariance matrices are the same. The test statistic used for testing homogeneous regression in the homogeneous innovation PMAR process is

$$(7.4) \quad \delta_{R|I} = \frac{|\hat{\Sigma}_I^{LG}|^{\frac{1}{2}(n-r)}}{|\hat{\Sigma}_{RI}^{LG}|^{\frac{1}{2}(n-r)}}$$

where $\hat{\Sigma}_I^{LG}$ is given by (5.3) and $\hat{\Sigma}_{RI}^{LG}$ is given by (5.4). We reject H_R for small values of $\delta_{R|I}$.

The asymptotic distribution of $-2\ln\delta_{R|I}$ is $\chi^2((\bar{p}d - p')m^2)$ under the assumption that the regression matrices $\Phi(k)$, $k=1, \dots, d$, do not depend on k .

Homogeneity in the PMAR process

To find a test statistic for testing time homogeneity in regression matrices as well as in innovation covariance, we look at a similar situation in the MLNM case, namely the simultaneous testing of equality of means and covariance matrices in several normal populations. The likelihood ratio test statistic for this case is the product of the test statistic for testing equality of means, given equality of covariances, and the test statistic for testing equality of covariances or, the product of the statistic for testing equality of covariances given equality of means and the test statistic for testing equality of means. The test statistic for the present situation is obtained by multiplying (7.1) with (7.3) or (7.2) with (7.4); this yields

$$\delta_{RI} = \delta_R \cdot \delta_I | R = \delta_I \cdot \delta_R | I = \frac{\prod_{k=1}^d |\hat{\Sigma}^{LG}_{(k)}|^{\frac{1}{2}(N_k - r_k^0)}}{|\hat{\Sigma}^{LG}_{RI}|^{\frac{1}{2}(n - r^0)}}$$

where $\hat{\Sigma}^{LG}$ is given by (4.4) and $\hat{\Sigma}^{LG}_{RI}$ is the LG estimate under time homogeneity of the parameters which is given by (5.4). Here $r^0 = \sum_{k=1}^d r_k^0$.

Under the assumption that the process is a aperiodic PMAR process (that is a MAR process), the asymptotic distribution of $-2\ln\delta_{RI}$ is $\chi^2((\bar{p}d - p')m^2 + \frac{1}{2}(d-1)m(m+1))$.

7.2. REDUCED CORRELATION STRUCTURE

The PMAR process reduces to a PMSAR process if the regression parameters are block diagonal. If in addition, in such a PMSAR process, the covariance matrices of the innovation process are block diagonal (with the same sizes of the blocks), the PMSAR process loses its (second order) covariational properties. In such a case it might be an advantage to analyse the (multivariate) subprocesses separately, for it reduces the complexity of the

calculations and the argument for analysing the processes jointly no longer exists, at least not on basis of second order properties.

7.2.1. Simple correlation in the PMAR process

To test simple correlation in the PMAR process we reject

$$H_S(k): \Phi_j(k) = B_{\underline{m}_S}(\Phi_j(k)), j=1, \dots, p_k$$

for small values of

$$(7.5) \quad \delta_S(k) = \frac{|\hat{\Sigma}^{LG}(k)|^{N_k - r_k^o}}{|\hat{\Sigma}_S^{LG}(k)|^{N_k - r_k^o}}$$

and

$$H_S = \bigcap_{k=1}^d H_S(k): \Phi_j(k) = B_{\underline{m}_S}(\Phi_j(k)), j=1, \dots, p_k, k=1, \dots, d$$

for small values of

$$\delta_S = \prod_{k=1}^d \delta_S(k)$$

Here $\hat{\Sigma}^{LG}(k)$ is the LG estimate given by (4.4) of $\Sigma(k)$ in the PMAR process and $\hat{\Sigma}_S^{LG}(k)$ is the corresponding estimate given by (5.7) in the PMSAR process. Under the assumption that the process is simple correlated, the statistics $-2\ln\delta_S(k)$ and $-2\ln\delta_S$ are asymptotically distributed as $\chi^2(2p_k(m^2 - \tilde{m}))$ and $\chi^2(2\bar{p}d(m^2 - \tilde{m}))$, respectively, where $\bar{p} = \sum_{k=1}^d p_k/d$ and $\tilde{m} = \sum_{i=1}^S m_i m_i^+$ for $m_i^+ = m_1 + \dots + m_i$ as before.

7.2.2. Block diagonal innovation covariances in the PMAR process

A problem in the multivariate normal regression model corresponding to this case is the testing of independence between subsets of variables.

The test that rejects the hypothesis

$$H_B(k): \Sigma(k) = B_{\underline{m}_s}(\Sigma(k)),$$

of block-diagonality of $\Sigma(k)$, for small values of

$$\delta'_B(k)^* = \frac{|\Sigma(k)^*|}{|\Sigma_{ii}(k)^*|}$$

or equivalent of

$$\delta_B(k)^* = \frac{|\Sigma(k)^*|^{N_k - r_k}}{|\Sigma_{ii}(k)^*|^{N_k - r_k}}$$

is (strictly) unbiased (Narain, 1950). The test that rejects

$$H_B = \bigcap_{k=1}^d H_B(k): \Sigma(k) = B_{\underline{m}_s}(\Sigma(k)), k=1, \dots, d$$

for small values of

$$\delta_B^* = \prod_{k=1}^d \delta_B(k)^*$$

is thus unbiased for H_B for all alternatives in the MLNM case.

To test H_B in the PMAR process, we use the test statistic

$$\delta_B = \prod_{k=1}^d \delta_B(k)$$

where

$$(7.6) \quad \delta_B(k) = \frac{|\hat{\Sigma}^{LG}_B(k)|^{N_k - r_k^o}}{|\hat{\Sigma}^{LG}_B(k)|^{N_k - r_k^o}}$$

and $\hat{\Sigma}^{LG}_B(k)$ is given by (5.8). Under the assumption that $\Sigma(k)$ is block diagonal, for $k=1, \dots, d$, the statistics $-2\ell n \delta_B(k)$ and $-2\ell n \delta_B$ are asymptotically distributed as $\chi^2(m^2 - m)$ and $\chi^2(d(m^2 - m))$, respectively.

The results of Corollary 6.1 can be used to construct a test statistic for testing the hypothesis

$$H_{ab}(k-i): \Sigma_{ab}(k-i) = 0$$

for some specified a, b, k and i . Since, for known regression matrices,

$\hat{S}_{ii}^{aa}(k)$ and $\hat{S}_{ii}^{bb}(k)$ are unbiased and consistent estimates of $\Sigma_{aa}(k-i)$

and $\Sigma_{bb}(k-i)$ respectively, by Corollary 6.1 (vi) we infer that under the

assumption that $H_{ab}(k-i)$ is true, the asymptotic distribution of

$$\begin{aligned} \tau_B^{ab}(k-i) &= (n/d) \hat{s}_{ii}^{ab}(k)^T (\hat{s}_{ii}^{aa}(k) \otimes \hat{s}_{ii}^{bb}(k))^{-1} \hat{s}_{ii}^{ab}(k) \\ (7.7) \quad &= (n/d) \text{tr}(\hat{s}_{ii}^{ba}(k) \hat{s}_{ii}^{aa}(k)^{-1} \hat{s}_{ii}^{ab}(k) \hat{s}_{ii}^{bb}(k)^{-1}) \end{aligned}$$

is $\chi^2(m_a m_b)$ as n tends to infinity. This test statistic corresponds to the Bartlett-Nanda-Pillai trace criterion statistic (Bartlett (1939), Nanda (1950), Pillai (1955)) for testing in the multivariate linear normal models. A similar test criterion has been suggested by Anderson (1978, Sec. 4.7) for testing that two innovation subprocesses are uncorrelated in a multivariate autoregressive scheme. The test statistic given by (7.7) is the analogue of that given by Anderson, modified for PMAR processes.

Several hypothesis can be combined into $H_B(k-i) = \bigcap_{a < b} H_{ab}(k-i)$ which represent the hypothesis that $\Sigma(k-i)$ is block diagonal. This hypothesis is rejected for large values of

$$\tau_B(k-i) = \sum_{a < b} \tau_B^{ab}(k-i).$$

The asymptotic distribution of $\tau_B(k-i)$ is $\chi^2(m^2 - m)$ if $H_B(k-i)$ is true, and thus the test is asymptotically equivalent to that based on $\delta_B(k-i)$.

7.2.3. Uncorrelated subprocesses in the PMAR process

Block-diagonal regression matrices in the PMBAR process

To test block-diagonality of the regression matrices in the PMBAR process, we reject $H_S(k)$ for small values of

$$(7.8) \quad \delta_{S|B}(k) = \frac{|\hat{\Sigma}_B^{LG}(k)|^{N_k - r_k}}{|\hat{\Sigma}_{SB}^{LG}(k)|^{N_k - r_k}}$$

and H_S for small values of

$$\delta_{S|B} = \prod_{k=1}^d \delta_{S|B}(k)$$

Here $\hat{\Sigma}_B^{LG}(k)$ is the LG estimate given by (5.8) of $\Sigma(k)$ in the case where

the innovation covariance matrices are block diagonal and $\hat{\Sigma}_{SB}^{LG}(k)$ is the corresponding estimate given by (5.9) when the process in addition is simple correlated. If $H_S(k)$ is true, $-2\ln\delta_{S|B}(k)$ is asymptotically distributed as $\chi^2(2p_k(m^2 - \tilde{m}))$ and if H_S is true, then the asymptotic distribution of $-2\ln\delta_{S|B}$ is $\chi^2(2d\bar{p}(m^2 - \tilde{m}))$.

Block-diagonal innovation covariances in the PMSAR process

If in the simple correlated model the covariance matrices $\Sigma(k)$, $k=1, \dots, d$ are block-diagonal (with block sizes identical to those of the regression matrices), the simple correlated structure collapses. If the process is Gaussian, the block-diagonal structure of $\Sigma(k)$, $k=1, \dots, d$, implies independence of the individual multivariate subprocesses.

To test block-diagonality of $\Sigma(k)$, $k=1, \dots, d$, we use the criterion to reject $H_B(k)$ for small values of

$$(7.9) \quad \delta_{B|S}(k) = \frac{|\hat{\Sigma}_S^{LG}(k)|^{N_k - r_k}}{|\hat{\Sigma}_{SB}^{LG}(k)|^{N_k - r_k}}$$

and H_B for small values of

$$\delta_{B|S} = \prod_{k=1}^d \delta_{B|S}(k)$$

where $\hat{\Sigma}_S^{LG}(k)$ is the LG estimate given by (5.7) of $\Sigma(k)$ in the simple correlated case and $\hat{\Sigma}_{SB}^{LG}(k)$ is the corresponding estimate given by (5.9) when $\Sigma(k)$ in addition is block-diagonal. Under the assumption that the innovation covariances are block-diagonal, the statistics $-2\ln\delta_{B|S}(k)$ and $-2\ln\delta_{B|S}$ are asymptotically distributed as $\chi^2(m^2 - \tilde{m})$ and $\chi^2(d(m^2 - \tilde{m}))$, respectively.

Block-diagonal process covariances in the PMAR process

When the regression matrices as well as the innovation covariance matrices are block-diagonal, then the subprocesses of the same dimension as the blocks are uncorrelated. The test statistic for testing such a non-correlation

- or independence in the Gaussian case - is obtained by multiplying (7.5) with (7.9) or (7.6) with (7.8).

The hypothesis

$$H_{SB}(k): \Sigma(k) = B_{\underline{m}_S}(\Sigma(k)) \quad \text{and} \quad \Phi_j(k) = B_{\underline{m}_S}(\Phi_j(k)), \quad j=1, \dots, p_k$$

is rejected for small values of

$$\delta_{SB}(k) = \delta_{B|S}(k) \cdot \delta_S(k) = \delta_{S|B}(k) \cdot \delta_B(k) = \frac{|\hat{\Sigma}^{LG}(k)|^{N_k - r_k}}{|\hat{\Sigma}_{SB}^{LG}(k)|^{N_k - r_k}}.$$

When this hypothesis is true, the asymptotic distribution of $-2\ln\delta_{SB}(k)$ is $\chi^2((2p_k+1)(m^2 - \tilde{m}))$. The hypothesis

$$H_{SB} = \bigcap_{k=1}^d H_{SB}(k)$$

which represents the hypothesis of uncorrelated subprocesses, is rejected for small values of

$$\delta_{SB} = \prod_{k=1}^d \delta_{SB}(k).$$

When the hypothesis H_{SB} is true, $-2\ln\delta_{SB}$ is asymptotically distributed as $\chi^2((2\bar{p}+1)d(m^2 - \tilde{m}))$.

8. PREDICTION

One major reason for adapting a statistical model to observed data is to use the estimated model for prediction purposes. In a PMAR model we distinguish between two types of predictions. Predicting the process one or more periods ahead will be referred to as long time prediction. Predicting the process in time points which are only small fractions of a whole period ahead will be called short time prediction. The reason for distinguishing between these types is that the approaches used for prediction are different.

8.1. SHORT TIME PREDICTION

The most natural approach is to use the estimated regression matrices

$\hat{\Phi}(k)$, $k=1, \dots, d$, to predict y_{n+1} with $\hat{y}_{n+1|n}$ given by

$$\hat{y}_{n+1|n} = \hat{\Phi}(n+1)y'_{n+1}$$

where $\hat{\Phi}(n+1) = \hat{\Phi}(1+(n \bmod d))$. By iterating this procedure we obtain

$$\hat{y}_{n+2|n} = \hat{\Phi}(n+2)\hat{y}'_{n+2|n}$$

where $\hat{y}'_{n+2|n} = \text{vec}(\hat{y}_{n+1|n}:y_n:\dots:y_{n+2-p_{n+2}})$ and $p_{n+2} = p_{1+(n+1 \bmod d)}$.

By repeated iteration we then obtain

$$\hat{y}_{n+k|n} = \hat{\Phi}(n+k)\hat{y}'_{n+k|n}$$

where $\hat{y}'_{n+k|n} = \text{vec}(\hat{y}_{n+k-1|n}:\hat{y}_{n+k-2|n}:\dots:\hat{y}_{n+k-p_{n+k}|n})$, and $\hat{y}_{n+k-i|n} = \hat{y}_{n+k-i|n}$ if $i < k$ and $\hat{y}_{n+k-i|n} = y_{n+k-i}$ if $i \geq k$ and where as before $p_{n+k} = p_{1+(n+k-1 \bmod d)}$.

8.2. LONG TIME PREDICTION

When the path of the process is to be predicted far ahead, it is possible to use the deperiodized process for predicting purposes. Analogously to

the short time predicting situation we use the estimated regression matrix

$\hat{\Psi} = (\hat{\Psi}_1 : \dots : \hat{\Psi}_p)$ to obtain an estimate $\hat{x}_{N+1|N}$ of x_{N+1} by

$$\hat{x}_{N+1|N} = \hat{\Psi} x'_{N+1}.$$

In companion form we have

$$\hat{x}'_{N+2|N} = \hat{\Psi} x'_{N+1}$$

where $\hat{x}'_{N+2|N} = \text{vec}(\hat{x}_{N+1|N} : x_N : \dots : x_{N+2-p})$ and

$$\hat{\Psi} = \begin{pmatrix} \hat{\Psi}_1 & \dots & \hat{\Psi}_{p-1} & \hat{\Psi}_p \\ I_{(p-1)md} & & & 0 \end{pmatrix}.$$

By iterating this procedure we then obtain

$$\hat{x}'_{N+k+1|N} = \hat{\Psi}^k x'_{N+1}$$

where for positive integers k ,

$$\hat{x}'_{N+k|N} = \text{vec}(\hat{x}_{N+k-1|N} : \hat{x}_{N+k-2|N} : \dots : \hat{x}_{N+k-p|N}),$$

and $\hat{x}_{N+k-i|N} = \hat{x}_{N+k-i|N}$ if $i < k$ and $\hat{x}_{N+k-i|N} = x_{N+k-i}$ if $i \geq k$.

REFERENCES

- Anděl, J. (1971). On the multiple autoregressive series. *Ann. Math. Statist.* 42, 755-759.
- Anděl, J. (1976). Autoregressive series with random parameter. *Math. Operationsforsch. u. Statist.*, 7, 735-741.
- Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. John Wiley and Sons, Inc. New York.
- Anderson, T.W. (1978). Repeated measurements on autoregressive processes. *J. Amer. Statist. Assoc.* 73, 371-378.
- Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*. 2nd Ed. John Wiley and Sons, Inc. New York.
- Anderson, T.W. and Mentz, R.P. (1980). On the structure of the likelihood function of autoregressive and moving average models. *J. Time Series analysis* 1, 83-94.
- Bartlett, M.S. (1933). On the theory of statistical regression. *Proc. Roy. Soc. Edinb.* 53, 260-283.
- Bartlett, M.S. (1939). A note on tests of significance in multivariate analysis. *Proc. Camb. Phil. Soc.* 35, 180-185.
- Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis. Forecasting and Control*. Holden-Day, Inc., California.
- Conlisk, J. (1974). Stability in a random coefficient model. *Int. Econ. Rev.* XV, 529-533.
- Conlisk, J. (1976). A further note on stability in a random coefficient model. *Int. Econ. Rev.* XVII, 759-764.
- Diananda, P.H. (1953). Some probability limit theorems with statistical applications. *Proc. Camb. Philos. Soc.* 49, 239-246.
- Dickey, J.M. (1967). Matricvariate generalizations of the multivariate t distribution and the inverted multivariate t distribution. *Ann. Math. Statist.* 38, 511-518.
- Durbin, J. (1959). Efficient estimation of parameters in moving-average models. *Biometrika* 46, 306-316.
- Eaton, M.L. (1983). *Multivariate Statistics. A Vector Space Approach*. John Wiley and Sons, Inc., New York.
- Gladyšev, E.G. (1961). Periodically correlated random sequences. *Soviet Math.* 2, 385-388.

- Granger, C.W.J. and Andersen, A.P. (1978). An Introduction to Bilinear Time Series Models. Vandenhoeck and Ruprecht, Göttingen.
- Hannan, E.J. (1970). Multiple Time Series. John Wiley and Sons, Inc., New York.
- Holmquist, B. (1985a). The direct product permuting matrices. Linear and Multilinear Algebra 17, 117-141.
- Holmquist, B. (1985b). Moments and cumulants from generating functions of Hilbert space-valued random variables and an application to the Wishart distribution. Lund Inst. of Technology, Techn. Report, 1-86.
- Holmquist, B. (1985c). An unbiased likelihood ratio test of equality of the covariance matrices in several multivariate normal populations with partially specified means. Unpublished manuscript.
- Hosking, J.R.M. (1980). The multivariate portmanteau statistic. J. Amer. Statist. Assoc. 75, 602-608.
- James, A.T. (1954). Normal multivariate analysis and the orthogonal group. Ann. Math. Statist. 25, 40-75.
- Jones, D.A. (1978). Non-linear autoregressive processes. Proc. Roy. Soc. Lond. A, 360, 71-95.
- Jones, R.H. and Brelsford, W.M. (1967). Time series with periodic structure. Biometrika 54, 403-408.
- Kshirsagar, A.M. (1961). Some extensions of the multivariate t-distribution and the multivariate generalizations of the distribution of the regression coefficient. Proc. Camb. Philos. Soc. 57, 80-85.
- Ledolter, J. (1980). Recursive estimation and adaptive forecasting in ARIMA models with time varying coefficients. In Second Applied Time Series Symposium (edited by D.F. Findley). Academic Press, New York.
- Magnus J.R. and Neudecker, H. (1979). The commutation matrix: Some properties and applications. Ann. Statist. 7, 381-394.
- Mahalanobis, P.C., Bose, R.C. and Roy, S.N. (1937). Normalization of statistical variates and the use of rectangular coordinates in the theory of sampling distributions. Sankyā 3, 1-40.
- Muirhead, R.J. (1982). Aspects of Multivariate Statistical Theory. John Wiley and Sons, Inc., New York.
- Nanda, D.N. (1950). Distributions of the sums of roots of a determinantal equation under a certain condition. Ann. Math. Statist. 21, 432-439.

- Narain, R.D. (1950). On the completely unbiased character of tests of independence in multivariate normal systems. *Ann. Math. Statist.* 21, 293-298.
- Nicholls, D.F. and Quinn, B.G. (1982). *Random Coefficient Autoregressive Models: An Introduction*. Springer-Verlag, New York.
- Osaki, T. (1980). Non-linear time series models for non-linear vibrations. *J. Appl. Prob.*, 17, 84-93.
- Pagano, M. (1973). When is a autoregressive scheme stationary? *Comm. Statist.* 1, 533-544.
- Pagano, M. (1978). On periodic and multiple autoregression. *Ann. Statist.* 6, 1310-1317.
- Pillai, K.C.S. (1955). Some new test criteria in multivariate analysis. *Ann. Math. Statist.* 26, 117-121.
- Risager, F. (1981). Model checking of simple correlated autoregressive processes. *Scand J. Statist.* 8, 137-153.
- Siddiqui, M.M. (1958). On the inversion of the sample covariance matrix in a stationary autoregressive process. *Ann. Math. Statist.* 29, 585-588.
- Srivastava, M.S. and Khatri, C.G. (1979). *An Introduction to Multivariate Statistics*. North Holland, New York.
- Subba Rao, T. (1970). The fitting of non stationary time series models with time dependent parameter. *J. Roy. Statist. Soc. B*, 32, 312-322.
- Thomas, H.A. and Fiering, M.B. (1962). Mathematical synthesis of stream-flow sequences for the analysis of river bassin by simulation. Ch. 12 of *Design of Water Resources Systems*. Harvard Univ. Press, Cambridge, Massachusetts.
- Tong, H. (1978). On a threshold model. In *Pattern Recognition and Signal Processing* (edited by C.H. Chen). Sijthoff and Noordhoff Int. Pub., The Netherlands.
- Tong, H. and Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data. *J. Roy. Statist. Soc. B*, 42, 245-293.
- Troutman, B.M. (1979). Some results in periodic autoregression. *Biometrika* 66, 219-228.
- Vinograd, B. (1950). Canonical positive definite matrices under internal linear transformations. *Proc. Amer. Math. Soc.* 1, 159-161.

- Whittle, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix. *Biometrika* 50, 129-134.
- Wilks, S.S. (1932). Certain generalizations in the analysis of variance. *Biometrika* 24, 471-494.
- Wishart, J. and Bartlett, M.S. (1933). The generalised product moment distribution in a normal system. *Proc. Camb. Phil. Soc.* 29, 260-270.

APPENDIX A: Proof of Lemma 3.1

Partition L , D and W into

$$L = \begin{pmatrix} L_{11} & 0 \\ L_{\cdot 1} & L''_1 \end{pmatrix}, \quad D = \begin{pmatrix} D_{11} & 0 \\ 0 & D''_1 \end{pmatrix} \quad \text{and} \quad W = \begin{pmatrix} W_{11} & W_{1\cdot} \\ W_{\cdot 1} & W''_1 \end{pmatrix}.$$

In order that $LDL^T = W$ it holds that

$$(A.1) \quad W_{11} = D_{11},$$

$$(A.2) \quad W_{\cdot 1} = L_{\cdot 1} D_{11},$$

$$(A.3) \quad W''_1 = L_{\cdot 1} D_{11} L_{\cdot 1}^T + L''_1 D''_1 L''_1^T,$$

From these equations we infer that

$$D_{11} = W_{11},$$

$$L_{\cdot 1} = W_{\cdot 1} D_{11}^{-1},$$

and

$$L''_1 D''_1 L''_1^T = W''_1 - L_{\cdot 1} D_{11} L_{\cdot 1}^T,$$

the two first of which give the results (3.2) and (3.3) for $j=1$. By replacing W and LDL^T by $W''_1 - L_{\cdot 1} D_{11} L_{\cdot 1}^T$ and $L''_1 D''_1 L''_1^T$ respectively in (A.1), (A.2) and (A.3), we obtain in a recursive manner the results in (3.2) and (3.3) for $j=2,3,\dots,d$.

If W is p.d., then so is W_{11} and hence also D_{11} . Further, $L''_1 D''_1 L''_1^T$ is then also p.d. since $W''_1 - L_{\cdot 1} D_{11} L_{\cdot 1}^T$ is nonsingular. The uniqueness of the decomposition thus follows from the unique solutions of the unique set of equations (A.1), (A.2) and (A.3) relating the blocks of L and D to the blocks of W .

APPENDIX B: Distributions of some compositions and decompositions

Let

$$\Gamma_{\underline{m}_d}(\frac{1}{2}n) = \pi^{m(m-1)/4} \prod_{i=1}^d \Gamma(\frac{1}{2}(n-m_i^++1))^{m_i}$$

where $m_i^+ = \sum_{j=1}^i m_j$. This mapping can be called the multivariate gamma function corresponding to the partition $\underline{m}_d = (m_1, \dots, m_d)$ of m . The multivariate gamma function corresponding to the complete partition $\underline{1}_m = (1, \dots, 1)$ of m ,

$$\Gamma_{\underline{1}_m}(\frac{1}{2}n) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma(\frac{1}{2}(n-i+1))$$

is the ordinary multivariate gamma function and is usually denoted $\Gamma_m(\frac{1}{2}n)$. However, this latter notation will here be used for the multivariate gamma function corresponding to the partition of m into a single part, that is for $d=1$ and $m_1 = m$,

$$\Gamma_m(\frac{1}{2}n) = \pi^{m(m-1)/4} \Gamma(\frac{1}{2}(n-m+1))^m.$$

We introduce some further distributional notations.

Wishart distribution: $W = (w_{ij})$ (in $S_m^+ \in W_m(p, \Sigma)$ if the ch.f. at $T \in S_m^+$ is $|I_m - 2iT\Sigma|^{-\frac{1}{2}p}$ for positive integers p and the density for $p \geq m$ is given by

$$(2^{\frac{1}{2}mp} \Gamma_{\underline{1}_m}(\frac{1}{2}p))^{-1} |\Sigma|^{-\frac{1}{2}p} |W|^{\frac{1}{2}p} \exp(-\frac{1}{2}\text{tr}(\Sigma^{-1}W))$$

with respect to $(dW)/|W|^{\frac{1}{2}(m+1)} = (\prod_{i \leq j} dw_{ij})/|W|^{\frac{1}{2}(m+1)}$. □

Matric chi distribution: L' (in $L_{\underline{m}_d}^+ \in W_{\underline{m}_d}^{\frac{1}{2}}(p, \Sigma)$ if the density for $\Sigma > 0$ is given by

$$(2^{\frac{1}{2}(mp + \frac{1}{2}(m^2 - 3m - 2m))} \Gamma_{\underline{m}_d}(\frac{1}{2}p))^{-1} |\Sigma|^{\frac{1}{2}p} |L'|^p \exp(-\frac{1}{2}\text{tr}(L'\Sigma^{-1}L'^T))$$

with respect to $(dL') / \prod_{i=1}^d |L'_{ii}|^{m_i^+}$ where L'_{ii} is the i th diagonal block of L' and $\tilde{m} = \sum_{i=1}^d m_i m_i^+$. □

Remark B.1. The matric chi distribution is a matric variate generalization of the Rayleigh and the Maxwell distributions and their higher degrees analogies on the real line. For $\underline{m}_d = \underline{1}_m$ this is the distribution of V^T in the Bartlett decomposition VV^T of W , where $W \in W_m(p, \Sigma)$, and for $\underline{m}_d = m$ this is the distribution of S in the p.d. square root decomposition of W into SS , where S is positive definite. The family of chi distributions is a set of distributions for different kinds of square roots of a Wishart distributed matrix. $\square$

Matric t distribution type I: $S = (s_{ij})$ (in $R^{m,n}$) $\in t_{m,n}(p, M, \Sigma, \Gamma)$ if the density is given by

$$\frac{\Gamma(\frac{1}{2}(p+mn))}{(p\pi)^{\frac{1}{2}mn} \Gamma(\frac{1}{2}p)} |\Sigma|^{-\frac{1}{2}n} |\Gamma|^{-\frac{1}{2}m} \left(1 + \frac{1}{p} \text{tr}(\Sigma^{-1}(S-M)\Gamma^{-1}(S-M)^T)\right)^{-\frac{1}{2}(p+mn)}$$

with respect to $(dS) = \prod_i \prod_j ds_{ij}$. $\square$

Matric t distribution type II: $S = (s_{ij})$ (in $R^{m,n}$) $\in T_{m,n}(p, M, \Sigma, \Gamma)$ if the density is given by

$$\frac{\Gamma_{\frac{1}{m}}(\frac{1}{2}(p+n))}{(p\pi)^{\frac{1}{2}mn} \Gamma_{\frac{1}{m}}(\frac{1}{2}p)} |\Sigma|^{-\frac{1}{2}n} |\Gamma|^{-\frac{1}{2}m} \left|I_m + \frac{1}{p} \Sigma^{-1}(S-M)\Gamma^{-1}(S-M)^T\right|^{-\frac{1}{2}(p+n)}$$

with respect to $(dS) = \prod_i \prod_j ds_{ij}$. $\square$

Remark B.2. The matric t distribution type II is essentially that of Dickey (1967), although the matrix is normalized so that

$$T_{m,n}(p, M, \Sigma, \Gamma) \rightarrow N_{m,n}(M, \Sigma, \Gamma) \text{ as } p \rightarrow \infty. \quad \square$$

Remark B.3. The matric t distributions of type I and II are both generalizations of the elliptical (or multivariate) t distribution: $s = (s_i)$ (in R^m) $\in t_m(p, \mu, \Sigma)$ if the density is given by

$$\frac{\Gamma(\frac{1}{2}(p+m))}{(p\pi)^{\frac{1}{2}m} \Gamma(\frac{1}{2}p)} |\Sigma|^{-\frac{1}{2}} \left(1 + \frac{1}{p}(s-\mu)^T \Sigma^{-1}(s-\mu)\right)^{-\frac{1}{2}(p+m)}$$

$$= \frac{\Gamma(\frac{1}{2}(p+m))}{(p\pi)^{\frac{1}{2}m} \Gamma(\frac{1}{2}p)} |\Sigma|^{-\frac{1}{2}} |I_m + \frac{1}{p} \Sigma^{-1}(s-\mu)(s-\mu)^T|^{-\frac{1}{2}(p+m)}$$

with respect to $(ds) = \prod_i ds_i$. If $s \in t_m(p, \mu, \Sigma)$ then $s \in t_{m,1}(p, \mu, \Sigma, 1)$ and $s^T \in T_{1,m}(p, \mu^T, 1, \Sigma)$. □

Recall that the Stiefel manifold $V_{n,m}$ in $R^{n,m}$ of m -frames in R^n is the space of linear isometries from R^m to R^n , that is $V_{n,m} = \{X \in R^{n,m}, n \geq m, X^T X = I_m\}$.

Uniform distribution on the Stiefel manifold $V_{n,m}$: $U = (u_{ij})$ (in $R^{n,m}$) $\in U_{n,m}$ if the density is given by

$$\frac{\Gamma_{\frac{1}{2}n}}{2^m \pi^{\frac{1}{2}mn}}$$

with respect to the O_n -invariant measure $((U:U^O)^T dU) = \prod_{1 \leq i \leq m} \prod_{i < j \leq n} u_j^T du_i$ where $(u_1 : \dots : u_m : u_{m+1} : \dots : u_n) = (U:U^O) \in O_n$ (the orthogonal group on $R^{n,n}$).

The invariant measure $((U:U^O)^T dU)$ can be written $\prod_{i=1}^d ((U_i : (U_1 : \dots : U_i)^O)^T dU_i)$ the density with respect to this measure,

$$\Gamma_{\frac{1}{2}n} / (2^m \pi^{\frac{1}{2}mn}) = \prod_{i=1}^d \Gamma_{\frac{1}{2}n} / (2^{m_i} \pi^{\frac{1}{2}m_i(n-m_{i-1}^+)})$$

Since for fixed $(U_1 : \dots : U_{i-1})$, the frame $(U_i : \dots : U_d)$ is a $m_i + \dots + m_d - n - m_{i-1}^+$ frame in $R^{n-m_{i-1}^+}$ because of the orthogonality relations between the columns of U . The distribution of U_i conditioned on $(U_1 : \dots : U_{i-1})$ is thus uniform over the space of m_i -frames in $R^{n-m_{i-1}^+}$. Cf. James (1954, p. 59) for the special case $m_d = \frac{1}{2}n$.

The following result is known, cf. e.g. Eaton (1983, p. 330).

LEMMA B.1. Let $X|W \in N_{m,n}(M, W^{-1}p, \Gamma)$ and $W \in W_m(p, \Sigma)$. Then it holds that $X \in T_{m,n}(p, M, \Sigma^{-1}, \Gamma)$. If $X|W \in N_{m,n}(M, \Sigma, W^{-1}p)$ and $W \in W_n(p, \Gamma)$ then it holds that $X^T \in T_{n,m}(p, M^T, \Gamma^{-1}, \Sigma)$.

PROOF. For the first case, the density of X is given by

$$\int (2\pi)^{-\frac{1}{2}mn} |W/p|^{-\frac{1}{2}n} |\Gamma|^{-\frac{1}{2}m} \exp(-\frac{1}{2} \text{tr}(W(X-M)\Gamma^{-1}(X-M)^T)/p)$$

$$\begin{aligned}
 & (2^{\frac{1}{2}mp} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p))^{-1} |\Sigma|^{-\frac{1}{2}p} |W|^{\frac{1}{2}p} \exp(-\frac{1}{2}\text{tr}(\Sigma^{-1}W)) (dW) / |W|^{\frac{1}{2}(m+1)} \\
 &= \frac{\Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n))}{(p\pi)^{\frac{1}{2}mn} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p)} |\Gamma|^{-\frac{1}{2}m} |\Sigma|^{-\frac{1}{2}p} |\Sigma^{-1} + (X-M)\Gamma^{-1}(X-M)^T/p|^{-\frac{1}{2}(p+n)} \\
 & \int (2^{\frac{1}{2}m(p+n)} \Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n)))^{-1} |\Sigma^{-1} + (X-M)\Gamma^{-1}(X-M)^T/p|^{\frac{1}{2}(p+n)} \\
 & |W|^{\frac{1}{2}(p+n)} \exp(-\frac{1}{2}\text{tr}(W(\Sigma^{-1} + (X-M)\Gamma^{-1}(X-M)^T/p))) (dW) / |W|^{\frac{1}{2}(m+1)} \\
 &= \frac{\Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n))}{(p\pi)^{\frac{1}{2}mn} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p)} |\Gamma|^{-\frac{1}{2}m} |\Sigma|^{-\frac{1}{2}p} |\Sigma^{-1} + (X-M)\Gamma^{-1}(X-M)^T/p|^{-\frac{1}{2}(p+n)} \\
 &= \frac{\Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n))}{(p\pi)^{\frac{1}{2}mn} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p)} |\Gamma|^{-\frac{1}{2}m} |\Sigma|^{\frac{1}{2}n} |I_m + \Sigma(X-M)\Gamma^{-1}(X-M)^T/p|^{-\frac{1}{2}(p+n)}
 \end{aligned}$$

which proves the first statement. Since $X \in N_{m,n}(M, \Sigma, W^{-1}p)$ if and only if $X^T \in N_{n,m}(M^T, W^{-1}p, \Sigma)$ the second statement immediately follows from the first statement. $\square$

The following result has been shown by Dickey (1967) for $M = 0$ and $\Sigma = I_m$.

LEMMA B.2. Let $W \in W_m(p, \theta)$, $p \geq m$, $\theta > 0$ and $X \in N_{m,n}(M, \Sigma, \Gamma)$, $\Sigma > 0$, $\Gamma > 0$ be independent and let $S = M + \sqrt{p} W^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} (X - M)$ where $A^{\frac{1}{2}}$ is any measurable square root of A , that is $A^{\frac{1}{2}} A^{\frac{1}{2}} = A$. Then

$$S \in T_{m,n}(p, M, \theta^{-1}, \Gamma).$$

PROOF. The simultaneous density of X and W is

$$\begin{aligned}
 & (2\pi)^{-\frac{1}{2}mn} |\Sigma|^{-\frac{1}{2}n} |\Gamma|^{-\frac{1}{2}m} \exp(-\frac{1}{2}\text{tr}(\Sigma^{-1}(X-M)\Gamma^{-1}(X-M)^T)) \\
 & (2^{\frac{1}{2}mp} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p))^{-1} |\theta|^{-\frac{1}{2}p} |W|^{\frac{1}{2}p} \exp(-\frac{1}{2}\text{tr}(\theta^{-1}W)) (dX)(dW) / |W|^{\frac{1}{2}(m+1)}.
 \end{aligned}$$

The transformation $(X, W) \rightarrow (S, W)$ defined by $S = M + \sqrt{p} W^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} (X - M)$ has the Jacobian $|d_X \square S| = |\sqrt{p} W^{-\frac{1}{2}} \Sigma^{-\frac{1}{2}} \otimes I_n| = p^{\frac{1}{2}mn} |W|^{-\frac{1}{2}n} |\Sigma|^{-\frac{1}{2}n}$. Thus the joint density of S and W is

$$\begin{aligned}
 & (2p\pi)^{-\frac{1}{2}mn} |\Gamma|^{-\frac{1}{2}m} \exp(-\frac{1}{2}\text{tr}(W^{\frac{1}{2}}(S-M)\Gamma^{-1}(S-M)^T W^{\frac{1}{2}T}/p)) \\
 & (2^{\frac{1}{2}mp} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p))^{-1} |\Theta|^{-\frac{1}{2}p} |W|^{\frac{1}{2}(p+n)} \exp(-\frac{1}{2}\text{tr}(\Theta^{-1}W)) (dS)(dW) / |W|^{\frac{1}{2}(m+1)} \\
 & = \frac{\Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n))}{(p\pi)^{\frac{1}{2}mn} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p)} |\Gamma|^{-\frac{1}{2}m} |\Theta|^{-\frac{1}{2}p} |\Theta^{-1} + (S-M)\Gamma^{-1}(S-M)^T/p|^{-\frac{1}{2}(p+n)} \\
 & (2^{\frac{1}{2}m(p+n)} \Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n)))^{-1} |\Theta^{-1} + (S-M)\Gamma^{-1}(S-M)^T/p|^{\frac{1}{2}(p+n)} \\
 & |W|^{\frac{1}{2}(p+n)} \exp(-\frac{1}{2}\text{tr}((\Theta^{-1} + (S-M)\Gamma^{-1}(S-M)^T/p)W)) (dS)(dW) / |W|^{\frac{1}{2}(m+1)}
 \end{aligned}$$

where the last step is written so that when integrating out W , we obtain the density of S as

$$\begin{aligned}
 & \frac{\Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n))}{(p\pi)^{\frac{1}{2}mn} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p)} |\Gamma|^{-\frac{1}{2}m} |\Theta|^{-\frac{1}{2}p} |\Theta^{-1} + (S-M)\Gamma^{-1}(S-M)^T/p|^{-\frac{1}{2}(p+n)} (dS) \\
 & = \frac{\Gamma_{\frac{1}{-m}}(\frac{1}{2}(p+n))}{(p\pi)^{\frac{1}{2}mn} \Gamma_{\frac{1}{-m}}(\frac{1}{2}p)} |\Gamma|^{-\frac{1}{2}m} |\Theta|^{\frac{1}{2}n} |I_m + \Theta(S-M)\Gamma^{-1}(S-M)^T/p|^{-\frac{1}{2}(p+n)} (dS).
 \end{aligned}$$

This completes the proof of the lemma. □

LEMMA B.3. Let $U \in U_{n,m}$ and $W \in W_m(n, \Theta)$ be independent and let $C = UW^{\frac{1}{2}}$ where $W^{\frac{1}{2}}$ is any measurable square root of W , that is $W^{\frac{1}{2}T}W^{\frac{1}{2}} = W$. Then

$$C \in N_{n,m}(0, I_n, \Theta).$$

PROOF. The simultaneous density of U and W is

$$2^{-m} (2\pi)^{-\frac{1}{2}mn} |\Theta|^{-\frac{1}{2}n} |W|^{\frac{1}{2}n} \exp(-\frac{1}{2}\text{tr}(\Theta^{-1}W))$$

with respect to $((U:U^O)^T dU)(dW) / |W|^{\frac{1}{2}(m+1)}$, where U^O is any orthogonal complement to U such that $(U:U^O) \in O_n$. It has been shown by James (1954, (18)) that

$$(dC) = 2^{-m} |C^T C|^{\frac{1}{2}(n-m-1)} d(C^T C) ((U:U^O)^T dU)$$

and hence the density can be written

$$(2\pi)^{-\frac{1}{2}mn} |\Theta|^{-\frac{1}{2}n} |C^T C|^{\frac{1}{2}n} \exp(-\frac{1}{2}\text{tr}(\Theta^{-1}C^T C))$$

with respect to $(dC)/|C^T C|^{1/2n}$; this is the density of $N_{n,m}(0, I_n, \theta)$ with respect to the θ_n -invariant measure. This proves the lemma. $\square$

The following results concerning the distributions of marginal and partial random dispersion matrices of a Wishart distributed matrix are well known, cf. e.g. Muirhead (1982, Corollary 3.2.6 and Theorem 3.2.10).

LEMMA B.4. Let $W \in W_m(n, \theta)$ and partition W into

$$W = \begin{pmatrix} W_{11} & W_{1'} \\ W_{,1} & W_1'' \end{pmatrix}$$

where W_{11} is m_1 -square, $W_{,1} = W_{1'}^T$ is of order m_1' by m_1 where $m_1' = m - m_1$ and where W_1'' is m_1' -square. Also partition θ into θ_{11} , $\theta_{,1}$, θ_1'' and θ_1' , analogously to the partition of W . Then

- (i) $W_{11} \in W_{m_1}(n, \theta_{11})$,
- (ii) $W_1'' \in W_{m_1'}(n, \theta_1'')$,
- (iii) $W_{11} - W_{,1} W_1''^{-1} W_{,1} \in W_{m_1}(n - m_1', \theta_{11} - \theta_{,1} \theta_1''^{-1} \theta_{,1})$,
- (iv) $W_1'' - W_{,1} W_{11}^{-1} W_{,1} \in W_{m_1'}(n - m_1, \theta_1'' - \theta_{,1} \theta_{11}^{-1} \theta_{,1})$,
- (v) $W_{,1} | W_{11} \in N_{m_1', m_1}(\theta_{,1} \theta_{11}^{-1} W_{11}, \theta_1'' - \theta_{,1} \theta_{11}^{-1} \theta_{,1}, W_{11})$,
- (vi) $W_{,1} W_{11}^{-1} | W_{11} \in N_{m_1', m_1}(\theta_{,1} \theta_{11}^{-1}, \theta_1'' - \theta_{,1} \theta_{11}^{-1} \theta_{,1}, W_{11}^{-1})$,
- (vii) $W_{,1} | W_1'' \in N_{m_1', m_1}(\theta_{,1} \theta_1''^{-1} W_1'', \theta_{11} - \theta_{,1} \theta_1''^{-1} \theta_{,1}, W_1'')$,
- (viii) $W_{,1} W_1''^{-1} | W_1'' \in N_{m_1', m_1}(\theta_{,1} \theta_1''^{-1}, \theta_{11} - \theta_{,1} \theta_1''^{-1} \theta_{,1}, W_1''^{-1})$,
- (ix) $W_{,1} W_{11}^{-1} W_{,1} | W_{11} \in W_{m_1'}(m_1, \theta_1'' - \theta_{,1} \theta_{11}^{-1} \theta_{,1}, \theta_{,1} \theta_{11}^{-1} W_{11} \theta_{11}^{-1} \theta_{,1})$,
- (x) $W_{,1} W_1''^{-1} W_{,1} | W_1'' \in W_{m_1'}(m_1', \theta_{11} - \theta_{,1} \theta_1''^{-1} \theta_{,1}, \theta_{,1} \theta_1''^{-1} W_1'' \theta_1''^{-1} \theta_{,1})$,
- (xi) $W_{11}^{-1} W_{,1} \in T_{m_1, m_1'}(n, \theta_{11}^{-1} \theta_{,1}, \theta_{11}^{-1}/n, \theta_1'' - \theta_{,1} \theta_{11}^{-1} \theta_{,1})$,
- (xii) $W_1''^{-1} W_{,1} \in T_{m_1', m_1}(n, \theta_1''^{-1} \theta_{,1}, \theta_1''^{-1}/n, \theta_{11} - \theta_{,1} \theta_1''^{-1} \theta_{,1})$

Further, $(W_{11} : W_{,1})$ and $W_1'' - W_{,1} W_{11}^{-1} W_{,1}$ are stochastically independent. $\square$

Remark B.4. The result in Lemma B.4 (xi) is given by Kshirsagar (1961). □

Cholesky decomposition of Wishart matrices: Bartlett decomposition

The following result concerns the distributions of the elements of the lower triangular matrix V with positive diagonal elements, in the decomposition VV^T of W known as the Bartlett decomposition (Bartlett, 1933) or decomposition into rectangular coordinates (Mahalanobis, Bose and Roy, 1937).

LEMMA B.5. Let VV^T be the Bartlett decomposition of $W \in W_m(n, \theta)$ into a lower triangular matrix V with positive diagonal elements. Further, partition V into

$$V = \begin{pmatrix} V_{11} & 0 \\ V_{,1} & V''_1 \end{pmatrix}$$

where V_{11} is m_1 -square, $V_{,1}$ is of order m'_1 by m_1 , and V''_1 is m'_1 -square. Let SS^T be the corresponding decomposition of θ into a lower triangular matrix S with positive diagonal elements. Partition S into S_{11} , $S_{,1}$, and S''_1 , analogously to the partition of V . Then

- (i) $V_{11}V_{11}^T \in W_{m_1}(n, S_{11}S_{11}^T),$
- (ii) $V_{,1}V_{,1}^T + V''_1V''_1{}^T \in W_{m'_1}(n, S_{,1}S_{,1}^T + S''_1S''_1{}^T),$
- (iii) $V_{11}(I_{m_1} - V_{,1}^T(V_{,1}V_{,1}^T + V''_1V''_1{}^T)^{-1}V_{,1})V_{11}^T$
 $\in W_{m_1}(n-m'_1, S_{11}(I_{m_1} - S_{,1}^T(S_{,1}S_{,1}^T + S''_1S''_1{}^T)^{-1}S_{,1})S_{11}^T)$
- (iv) $V''_1V''_1{}^T \in W_{m'_1}(n-m_1, S''_1S''_1{}^T),$
- (v) $V_{,1}V_{11}^T | V_{11} \in N_{m'_1, m_1}(S_{,1}S_{11}^{-1}V_{11}V_{11}^T, S''_1S''_1{}^T, V_{11}V_{11}^T)$
- (vi) $V_{,1}V_{11}^{-1} | V_{11} \in N_{m'_1, m_1}(S_{,1}S_{11}^{-1}, S''_1S''_1{}^T, (V_{11}V_{11}^T)^{-1})$
- (vii) $V_{,1}V_{,1}^T | V_{11} \in W_{m'_1}(m_1, S''_1S''_1{}^T, S_{,1}S_{11}^{-1}V_{11}V_{11}^T S_{11}^{-T}S_{,1}^T)$
- (viii) $V_{,1} | V_{11} \in N_{m'_1, m_1}(S_{,1}S_{11}^{-1}V_{11}, S''_1S''_1{}^T, I_{m_1})$

$$(ix) \quad (V_{,1} V_{11}^{-1})^T \in T_{m_1, m_1} (n, (S_{,1} S_{11}^{-1})^T, (S_{11} S_{11}^T)^{-1} / n, S_1'' S_1''^T)$$

$$(x) \quad (V_{,1} V_{,1}^T + V_1'' V_1''^T)^{-1} V_{,1} V_{11}^T \in T_{m_1, m_1} (n, (S_{,1} S_{,1}^T + S_1'' S_1''^T)^{-1} S_{,1} S_{11}^T, \\ (S_{,1} S_{,1}^T + S_1'' S_1''^T)^{-1} / n, S_{11} (I + S_{,1}^T (S_1'' S_1''^T)^{-1} S_{,1})^{-1} S_{11}^T)$$

Further $(V_{11} : V_{,1}^T)$ and V_1'' are stochastically independent.

PROOF. Since the following relations hold

$$(B.1i) \quad W_{11} = V_{11} V_{11}^T$$

$$(B.1ii) \quad W_1'' = V_{,1} V_{,1}^T + V_1'' V_1''^T$$

$$(B.1iii) \quad W_{11} - W_{,1} W_1''^{-1} W_{,1} = V_{11} (I_{m_1} - V_{,1}^T (V_{,1} V_{,1}^T + V_1'' V_1''^T)^{-1} V_{,1}) V_{11}^T$$

$$(B.1iv) \quad W_1'' - W_{,1} W_{11}^{-1} W_{,1} = V_1'' V_1''^T$$

$$(B.1v) \quad W_{,1} = V_{,1} V_{11}^T$$

$$(B.1vi) \quad W_{,1} W_{11}^{-1} = V_{,1} V_{11}^{-1}$$

$$(B.1vii) \quad W_{,1} W_{11}^{-1} W_{,1} = V_{,1} V_{,1}^T$$

$$(B.1viii) \quad W_{,1} W_1''^{-1} W_{,1} = V_{11} V_{,1}^T (V_{,1} V_{,1}^T + V_1'' V_1''^T)^{-1} V_{,1} V_{11}^T$$

and also the corresponding relations obtained by the substitution (W, V)

$\rightarrow (\Theta, S)$, the results of Lemma B.5 follow from Lemma B.4. □

Remark B.5. If $S_{,1} = 0$ (or equivalent $\Theta_{,1} = 0$), it follows from Lemma B.5 (x) that

$$V_{,1} \in N_{m_1, m_1} (0, S_1'' S_1''^T, I_{m_1})$$

independently of V_{11} . Then the result in Lemma B.5 (xi) (for $S_{,1} = 0$)

follows from Lemma B.2. □

Block modified Cholesky decomposition of Wishart matrices

THEOREM B.1. Let $W \in W_m(n, \Theta)$ and let the block modified Cholesky decompositions of W and Θ be $W = L(\bigoplus_{i=1}^s D_{ii})L^T$ and $\Theta = \Lambda(\bigoplus_{i=1}^s \Delta_{ii})\Lambda^T$ with $L = (L_{ij})$ and $\Lambda = (\Lambda_{ij})$, $i, j=1, \dots, s$, where the m_i by m_j matrices L_{ij} and Λ_{ij} are equal to zero matrices if $i < j$ and I_{m_i} if $i=j$ and

D_{ii} and Λ_{ii} are m_i -square p.d. matrices. Let $m'_k = m - m_1 - m_2 - \dots - m_k$,
 $m'_0 = m$,

$$L''_{k-1} = \begin{pmatrix} L_{kk} & \dots & L_{ks} \\ \vdots & & \vdots \\ L_{sk} & \dots & L_{ss} \end{pmatrix}, \quad k=2, \dots, s, \quad L_{,k} = \begin{pmatrix} L_{k+1,k} \\ \vdots \\ L_{sk} \end{pmatrix}$$

and $D''_k = \bigoplus_{j=k+1}^s D_{jj}$ for $k=1, \dots, s-1$. Then

- (i) $D_{jj} \in W_{m_j}(n-m+m'_{j-1}, \Delta_{jj})$, $j=1, \dots, s$
- (ii) $L_{,1} D_{11} L_{,1}^T + L_1'' D_1'' L_1''^T \in W_{m_1}(n, \Lambda_{,1} \Delta_{11} \Lambda_{,1}^T + \Lambda_1'' \Delta_1'' \Lambda_1''^T)$
- (iii) $(D_{11}^{-1} + L_{,1}^T (L_1'' D_1'' L_1''^T)^{-1} L_{,1})^{-1} \in W_{m_1}(n-m'_1, (\Delta_{11}^{-1} + \Lambda_{,1}^T (\Lambda_1'' \Delta_1'' \Lambda_1''^T)^{-1} \Lambda_{,1})^{-1})$
- (iv) $L_j'' D_j'' L_j''^T \in W_{m_j}(n-m+m'_j, \Lambda_j'' \Delta_j'' \Lambda_j''^T)$, $j=1, \dots, s$,
- (v) $L_{,j} D_{jj} | D_{jj} \in N_{m'_j, m_j}(\Lambda_{,j} D_{jj}, \Lambda_j'' \Delta_j'' \Lambda_j''^T, D_{jj})$,
- (vi) $L_{,j} | D_{jj} \in N_{m'_j, m_j}(\Lambda_{,j}, \Lambda_j'' \Delta_j'' \Lambda_j''^T, D_{jj}^{-1})$, $j=1, \dots, s-1$,
- (vii) $D_{11} L_{,1}^T | (L_{,1} D_{11} L_{,1}^T + L_1'' D_1'' L_1''^T) \in N_{m_1, m'_1}(\Delta_{11} \Lambda_{,1}^T (L_{,1} D_{11} L_{,1}^T + L_1'' D_1'' L_1''^T),$
 $(\Delta_{11}^{-1} + \Lambda_{,1}^T (\Lambda_1'' \Delta_1'' \Lambda_1''^T)^{-1} \Lambda_{,1})^{-1}, L_{,1} D_{11} L_{,1}^T + L_1'' D_1'' L_1''^T)$
- (viii) $L_{,1} D_{11} L_{,1}^T | D_{11} \in W_{m_1}(m_1, \Lambda_1'' \Delta_1'' \Lambda_1''^T, \Lambda_{,1} D_{11} \Lambda_{,1}^T)$

and

- (ix) $L_{,j}^T \in {}^{-T}N_{m_j, m'_j}(n-m+m'_{j-1}, \Lambda_{,j}^T, \Delta_{jj}^{-1}/n, \Lambda_j'' \Delta_j'' \Lambda_j''^T)$, $j=1, \dots, s-1$,

and the r.v.'s $(D_{jj} : L_{,j}^T)$, $j=1, \dots, s-1$ and D_{ss} are all stochastically independent.

PROOF. Let

$$L = \begin{pmatrix} L_{11} & 0 \\ L_{,1} & L_1'' \end{pmatrix}, \quad D = \begin{pmatrix} D_{11} & 0 \\ 0 & D_1'' \end{pmatrix}$$

and

$$\Lambda = \begin{pmatrix} \Lambda_{11} & 0 \\ \Lambda_{,1} & \Lambda_1'' \end{pmatrix}, \quad \Delta = \begin{pmatrix} \Delta_{11} & 0 \\ 0 & \Delta_1'' \end{pmatrix}.$$

Then we have

$$(B.2i) \quad W_{11} = D_{11}$$

$$(B.2ii) \quad W_1'' = L_{,1} D_{11} L_{,1}^T + L_1'' D_1'' L_1''^T$$

$$(B.2iii) \quad W_{11} - W_1 W_1''^{-1} W_{,1} = D_{11} - D_{11} L_{,1}^T (L_{,1} D_{11} L_{,1}^T + L_1'' D_1'' L_1''^T)^{-1} L_{,1} D_{11} \\ = (D_{11}^{-1} + L_{,1}^T (L_1'' D_1'' L_1''^T)^{-1} L_{,1})^{-1}$$

$$(B.2iv) \quad W_1'' - W_{,1} W_{11}^{-1} W_{,1}' = L_1'' D_1'' L_1''^T$$

$$(B.2v) \quad W_{,1} = L_{,1} D_{11}$$

$$(B.2vi) \quad W_{,1} W_{11}^{-1} = L_{,1}$$

$$(B.2vii) \quad W_{,1} W_{11}^{-1} W_{,1}' = L_{,1} D_{11} L_{,1}^T$$

and also the corresponding relations obtained by the substitution (D, L, V)

$\rightarrow (\Delta, \Lambda, S)$. From Lemma B.4 (i), (iv), (vi) and (xi) we then obtain

$$D_{11} \in W_{m_1}(n, \Delta_{11}),$$

$$(B.3) \quad L_1'' D_1'' L_1''^T \in W_{m_1}(n-m_1, \Lambda_1'' \Delta_1'' \Lambda_1''^T),$$

$$L_{,1} | D_{11} \in N_{m_1', m_1}(\Lambda_{,1}, \Lambda_1'' \Delta_1'' \Lambda_1''^T, D_{11}^{-1}),$$

and

$$L_{,1}^T \in T_{m_1, m_1'}(n, \Lambda_{,1}^T, \Delta_{11}^{-1}/n, \Lambda_1'' \Delta_1'' \Lambda_1''^T)$$

which proves the results in (i), (iv), (vi) and (xi) for $j=1$. We also infer from Lemma B.4 that $(D_{11} : L_{,1}^T)$ and $(D_1'' : L_1'')$ are stochastically independent. By using this result and the result in (B.3) recursively, we obtain the distributions of D_{jj} , $L_{,j} | D_{jj}$ and $L_{,j}$ for $j=2, \dots, s-1$ and D_{ss} . This proves the lemma. □

Gram-Schmidt triangular decomposition of Gaussian matrices

The following result is known, cf. Eaton (1983, Proposition 7.2 and Example 7.1).

LEMMA B.6. Let UV^T be the Gram-Schmidt decomposition of the matrix

$X \in N_{n,m}(0, I_n, \Theta)$, where $n \geq m$ and $\Theta > 0$, into a n by m linear isometry

U and a lower triangular m -square matrix V . Then U and V are stochas-

tically independent,

$$U \in U_{n,m}$$

and V has the distribution given in Lemma B.5. □

Remark B.6. The Gram-Schmidt triangular decomposition of a Gaussian matrix $X \in N_{n,m}(0, I_n, \theta)$ into UV^T corresponds to the Cholesky decomposition of Wishart matrices (the Bartlett decomposition) in the sense that if X is decomposed into UV^T then $W \equiv X^T X$ is decomposed into $VU^T UV^T = VV^T$ since $U^T U = I_m$. In Theorem B.2 below we give a decomposition of a Gaussian matrix which corresponds to the block modified Cholesky decomposition of the corresponding Wishart matrix. □

Block modified Gram-Schmidt triangular decomposition of Gaussian matrices

We need the following generalization of the Gram-Schmidt orthogonalization procedure:

LEMMA B.7. (Block Gram-Schmidt suborthogonalization) Let $X = (X_1 : \dots : X_d)$ be a n by m matrix with n by m_i blocks X_i , $i=1, \dots, d$, where $m_1 + \dots + m_d = m$. Define, in recursive manner, n by m_i matrices E_i , $i=1, \dots, d$, by

$$(B.4) \quad E_i = X_i - \sum_{j=1}^{i-1} E_j (E_j^T E_j)^- E_j^T X_i.$$

for some g -inverse $(E_j^T E_j)^-$ of $E_j^T E_j$. Then it holds that

$$E_j^T E_i = \begin{cases} 0 & i \neq j \\ E_i^T X_i & i = j. \end{cases}$$

and

$$E_i^T X_j = 0$$

for $i > j$.

PROOF. Assume that $E_k^T E_j = 0$ for $j, k=1, \dots, i-1$, $k \neq j$ and $E_k^T E_k = E_k^T X_k$

for $k=1, \dots, i-1$. Then we have for $k=1, \dots, i-1$,

$$E_k^T E_i = E_k^T X_i - \sum_{j=1}^{i-1} E_k^T E_j (E_j^T E_j)^{-1} E_j^T X_i = E_k^T X_i - E_k^T X_i = 0$$

and

$$E_i^T E_i = E_i^T X_i - \sum_{j=1}^{i-1} E_i^T E_j (E_j^T E_j)^{-1} E_j^T X_i = E_i^T X_i.$$

Since it holds that $E_1^T E_2 = E_1^T X_2 - E_1^T E_1 (E_1^T E_1)^{-1} E_1^T X_2 = 0$ and $E_1^T E_1 = E_1^T X_1$, the first result follows by an induction argument.

To show the second part, we write $EL^T = X$ where $E = (E_1 : \dots : E_d)$ and the the block unit lower triangular matrix $L = (L_{ij})$ is given by (B.5) below. Hence

$$E^T E = E^T X L^{-T}.$$

Since L^{-T} is block unit upper triangular and $E^T E$ is block diagonal, it follows by identification of blocks in a recursive way that the subdiagonal blocks of $E^T X$ are all zero blocks. This proves the second part. $\square$

By means of this block suborthogonalization procedure, we can show the following block modified Gram-Schmidt triangular decomposition of matrices.

THEOREM B.2. Let $X = (X_1 : \dots : X_d)$ be a n by m matrix with n by m_i blocks X_i , $i=1, \dots, d$, where $m_1 + \dots + m_d = m$. Then there exist a decomposition into a n by m linear isometry U , a block unit lower triangular matrix $L = (L_{ij})$ with blocks L_{ij} of size m_i by m_j and a block diagonal matrix $D^{\frac{1}{2}} = \bigoplus_{i=1}^d D_{ii}^{\frac{1}{2}}$ with block $D_{ii}^{\frac{1}{2}}$ of size m_i -square, such that

$$X = U D^{\frac{1}{2}} L^T.$$

The blocks of $D \equiv D^{\frac{1}{2}T} D^{\frac{1}{2}} = \bigoplus_{i=1}^d D_{ii}$ and L are given by

$$D_{ii} = X_i^T (I_n - \sum_{j=1}^{i-1} E_j D_{jj}^{-1} E_j^T) X_i = E_i^T E_i$$

and

$$(B.5) \quad L_{ij} = \begin{cases} I_{m_i} & i=j \\ 0 & i < j \\ X_i^T E_j (E_j^T E_j)^{-1} & i > j. \end{cases}$$

where E_j is defined in (B.4).

If X has rank m ($n \geq m$), then L and $UD^{\frac{1}{2}}$ are unique. A particular choice of U and $D^{\frac{1}{2}}$ is $D^{\frac{1}{2}} = (E^T E)^{\frac{1}{2}}$ and $U = E(D^{\frac{1}{2}})^{-1}$ for a square root $(E^T E)^{\frac{1}{2}}$ of $E^T E$, where $E = (E_1 : \dots : E_d)$.

PROOF. By using Lemma B.7 we can write $EL^T = X$, where $L = (L_{ij})$ is given by (B.5). From Lemma B.7 it follows that $E^T E$ is blockdiagonal with the i th diagonal block equal to $E_i^T E_i = D_{ii}$. Letting $D^{\frac{1}{2}} = \bigoplus_{i=1}^d D_{ii}^{\frac{1}{2}}$ be such that $D^{\frac{1}{2}T} D^{\frac{1}{2}} = D \equiv E^T E$, it follows by a result of Vinograd (1950) that $E = UD^{\frac{1}{2}}$ for some n by m linear isometry U .

If X is of rank m , then so is E since L is nonsingular. Thus if $EL^T = FM^T$, where $E^T E = D_1$ and $F^T F = D_2$ are both blockdiagonal, then $X^T X = LD_1 L^T = MD_2 M^T$ and thus $D_1 = KD_2 K^T$ where $K = L^{-1}M$. Since D_2 is nonsingular, $K = I$ and thus $L = M$, $D_1 = D_2$ and $E = F$.

Since for any choice of the square root, the particular choice $D^{\frac{1}{2}} = (E^T E)^{\frac{1}{2}}$ and $U = E(D^{\frac{1}{2}})^{-1}$ satisfy $UD^{\frac{1}{2}} = E$ and $U^T U = (D^{\frac{1}{2}})^{-T} E^T E (D^{\frac{1}{2}})^{-1} = I_m$; this choice is a solution of the decomposition of E into $UD^{\frac{1}{2}}$. $\square$

Remark B.7. For the case $\frac{m}{d} = \frac{1}{m}$, the decomposition into a linear isometry and a triangular matrix is well known. $\square$

Remark B.8. Sometimes it is convenient to combine $D^{\frac{1}{2}}$ and L into a block lower triangular matrix $V = LD^{\frac{1}{2}T}$. In this case the decomposition $X = UV^T$ is referred to as a block Gram-Schmidt triangular decomposition of X . In other cases it is convenient to combine U and $D^{\frac{1}{2}}$ into a block-diagonal-square-root $C \equiv UD^{\frac{1}{2}}$ ($C^T C = D^{\frac{1}{2}T} D^{\frac{1}{2}} \equiv D$ (blockdiagonal)) giving the decomposition $X = CL^T$. $\square$

Using Theorem B.2 we can show the following result for the factor

matrices in the block modified Gram-Schmidt triangular decomposition of Gaussian matrices.

THEOREM B.3. Let $X \in N_{n,m}(0, I_n, \Theta)$, where $n \geq m$, and let the block modified Gram-Schmidt triangular decomposition of X be $X = UD^{\frac{1}{2}}L^T$ into $U \in V_{n,m}$, $D^{\frac{1}{2}} = \bigoplus_{i=1}^d D_{ii}^{\frac{1}{2}}$ where $D_{ii}^{\frac{1}{2}} \in L_{m_i}^{+,+}$ and $L^T \in L_{\frac{m}{d}}^{+,I}$. Then U and $(D^{\frac{1}{2}}:L^T)$ are stochastically independent,

$$(i) \quad U \in U_{n,m}$$

$$(ii) \quad D^{\frac{1}{2}}L^T \in W_{\frac{m}{d}}^{\frac{1}{2}}(n, \Theta)$$

(iii) The $D_{ii}^{\frac{1}{2}}$'s are stochastically independent and the matrices

$$D_{ii}^{\frac{1}{2}} \in W_{m_i}^{\frac{1}{2}}(n-m_{i-1}, \Delta_{ii}) \quad (i=1, \dots, d)$$

$$(iv) \quad UD^{\frac{1}{2}} \in N_{n,m}(0, I_n, + \bigoplus_{i=1}^d \Delta_{ii})$$

and L have the distributions given in Theorem B.1.

PROOF. (i) For $n \geq m$ the rank of X is m with probability one. Let $U(X)$ and $L'(X)$ be the unique elements $U \in V_{n,m}$ and $L' \equiv D^{\frac{1}{2}}L^T \in L_{\frac{m}{d}}^{+,+}$ respectively, given by Theorem B.2, such that $X = UL'$. For any $G \in O_n$ we have that $U(GX) = GU(X)$. Hence, for any Borel set B of $V_{n,m}$, it follows that

$$P(U(X) \in B) = P(U(GX) \in B) = P(GU(X) \in B) = P(U(X) \in G^TB);$$

the second equality follows from the fact that GX has the same distribution as X . The distribution of $U(X)$ is thus O_n -invariant and hence equal to the unique probability measure on $V_{n,m}$, that is $U_{n,m}$. This proves (i).

(ii) Let

$$f_0(X^TX) = (2\pi)^{-\frac{1}{2}mn} |\Theta|^{-\frac{1}{2}n} |X^TX|^{\frac{1}{2}n} \exp(-\frac{1}{2}\text{tr}(X^TX\Theta^{-1})),$$

$$\mu(dX) = dX/|X^TX|^{\frac{1}{2}n},$$

$$g_0(L') = (2^{\frac{1}{2}(mn+\frac{1}{2}(m^2-3m-2m))}) \Gamma_{\frac{m}{d}}(\frac{1}{2}n)^{-1}$$

$$|\Theta|^{-\frac{1}{2}n} |L'|^n \exp(-\frac{1}{2}\text{tr}(L'\Theta^{-1}L'^T))$$

and

$$v_r(dL') = dL' / \prod_{i=1}^d |L'_{ii}|^{m_i^+}.$$

We shall show that

$$\int_{R^{m,n}} h(L'(X)) f_0(X^T X) \mu(dX) = \int_{L'_{\frac{m}{d}}^+} h(L') g_0(L') v_r(dL')$$

for all integrable functions h on $L'_{\frac{m}{d}}^+$.

Let $G \in O_n$ and $L \in L_{\frac{m}{d}}^+$, then $L'((G, L)X) = L'(GXL^T) = L'(GUL^T L^T) = L'L^T$. Hence $L'(\cdot)$ is equivariant with $L'((G, L)X) = \overline{(G, L)}L'(X)$ where $\overline{(G, L)}L' = L'L^T$. For $K \subset L'_{\frac{m}{d}}^+$, the set $L'^{-1}(K)$ is compact and thus $\mu(L'^{-1}(K)) < +\infty$. It thus follows from the equivariational property of $L'(\cdot)$ that the measure v defined by $v(\cdot) = \mu \circ L'^{-1}(\cdot)$, that is $v(L') = \mu(L'^{-1}(L')) = \mu(X)$, is a right invariant measure for actions from the group over $L'_{\frac{m}{d}}^+$ and it holds that

$$(B.6) \quad \int_{R^{n,m}} f(L'(X)) \mu(dX) = \int_{L'_{\frac{m}{d}}^+} f(L') v(dL')$$

Let L' and R be both in $L'_{\frac{m}{d}}^+$ and define g on $L'_{\frac{m}{d}}^+$ by $g(L') = L'R$. Then the Jacobian of g at R is

$$\prod_{i=1}^d |R_{ii}|^{m_i^+},$$

where $m_i^+ = m_1 + \dots + m_i$, and R_{ii} is the i th diagonal block of R and of size m_i -square. It thus follows that

$$v_r(dL') = dL' / \prod_{i=1}^d |L'_{ii}|^{m_i^+}$$

is a right invariant measure on $L'_{\frac{m}{d}}^+$.

To determine the right invariant Haar measure v corresponding to μ through $v(\cdot) = \mu \circ L'^{-1}(\cdot)$ — which is a constant c times v_r — we choose

$$f(L') = (2\pi)^{-\frac{1}{2}mn} |L'^T L'|^{-\frac{1}{2}n} \exp(-\frac{1}{2} \text{tr}(L'^T L'))$$

in (B.6). We then have

$$1 = c \int_{L'_{\frac{m}{d}}^+} (2\pi)^{-\frac{1}{2}nm} |L'^T L'|^{-\frac{1}{2}n} \exp(-\frac{1}{2} \text{tr}(L'^T L')) dL' / \prod_{i=1}^d |L'_{ii}|^{m_i^+}$$

$$= c \int_{L', +} (2\pi)^{-\frac{1}{2}nm} \prod_{i=1}^d |L'_{ii}|^{n-m_i^+} \exp(-\frac{1}{2} \sum_{i \leq j} \text{tr}(L'_{ij}^T L'_{ij})) (dL')$$

where L'_{ij} is the (i,j) th block of size m_i by m_j of L' . Letting $S_m^{\frac{1}{2}+} = \{X \in R^{m,m}, X^T X > 0\}$ we factorize the terms on the right hand side into $c(2\pi)^{-\frac{1}{2}nm}$ times the product of the integrals

$$(B.7) \quad \int_{S_{m_i}^{\frac{1}{2}+}} |L'_{ii}|^{n-m_i^+} \exp(-\frac{1}{2} \text{tr}(L'_{ii}^T L'_{ii})) (dL'_{ii}),$$

$i=1, \dots, d$, and

$$\int_{R^{m_i, m_j}} \exp(-\frac{1}{2} \text{tr}(L'_{ij}^T L'_{ij})) (dL'_{ij}),$$

$i, j=1, \dots, d, i < j$. For $i < j$ it follows from normal theory that

$$\int_{R^{m_i, m_j}} \exp(-\frac{1}{2} \text{tr}(L'_{ij}^T L'_{ij})) (dL'_{ij}) = (2\pi)^{\frac{1}{2}m_i m_j}.$$

By an application of a result of Vinograd (1950), we have that if $W_1^{\frac{1}{2}}$ and $W_2^{\frac{1}{2}}$ are two square roots of W , that is $W_1^{\frac{1}{2}T} W_1^{\frac{1}{2}} = W_2^{\frac{1}{2}T} W_2^{\frac{1}{2}} = W$, then there exist an orthogonal matrix G such that $GW_1^{\frac{1}{2}} = W_2^{\frac{1}{2}}$. Since (B.7) is invariant under orthogonal transformations, we may thus for a moment assume that $L'_{ii} \in L_{-m_i}^{\frac{1}{2}+}$ is the triangular square root of $L'_{ii}^T L'_{ii}$. With λ_{jk} denoting the (j,k) 'th element of L'_{ii} this yields

$$(B.8) \quad \int_{S_{m_i}^{\frac{1}{2}+}} |L'_{ii}|^{n-m_i^+} \exp(-\frac{1}{2} \text{tr}(L'_{ii}^T L'_{ii})) (dL'_{ii}) = \int \dots \int \prod_{j=1}^{m_i} \lambda_{jj}^{n-m_i^+} \exp(-\frac{1}{2} \sum_{j \leq k} \lambda_{jk}^2) \prod_{j \leq k} d\lambda_{jk}$$

where λ_{jj} ranges from 0 to ∞ and λ_{jk} for $j < k$, from $-\infty$ to ∞ . By using normal theory and the relation

$$\int_{x>0} x^r \exp(-\frac{1}{2}x^2) dx = 2^{\frac{1}{2}(r-1)} \Gamma(\frac{1}{2}(r+1)),$$

it follows that (B.8) is equal to

$$2^{\frac{1}{2}(n-m_i^+-1)m_i} \Gamma(\frac{1}{2}(n-m_i^++1))^{m_i} (2\pi)^{m_i(m_i-1)/4}.$$

Some calculation thus yields that

$$c = (2\pi)^{\frac{1}{2}mn} (2^{\frac{1}{2}(mn+\frac{1}{2}(m^2-3m-2\tilde{m}))}) \Gamma_{\frac{m}{d}}(\frac{1}{2}n))^{-1}.$$

where $\tilde{m} = \sum_{i=1}^d m_i m_i^+$. Hence from (B.6),

$$\int_{R^{n,m}} h(L'(X)) f_0(X^T X) \mu(dX) = \int_{L'^+, \frac{m}{d}} h(L') f_0(L'^T L') c \nu_r(dL')$$

This shows that the density of L' with respect to ν_r is

$$\begin{aligned} g_0(L') &= c f_0(L'^T L') \\ &= (2^{\frac{1}{2}(mn+\frac{1}{2}(m^2-3m-2\tilde{m}))}) \Gamma_{\frac{m}{d}}(\frac{1}{2}n))^{-1} |\Theta|^{-\frac{1}{2}n} |L'|^n \exp(-\frac{1}{2}\text{tr}(L' \Theta^{-1} L'^T)) \end{aligned}$$

which proves (ii).

(iii) Since we have $D_{ii}^{\frac{1}{2}} \equiv L'_{ii}$, we calculate the marginal simultaneous distribution of the diagonal blocks of L' . Let $\Lambda(\bigoplus_{i=1}^d \Delta_{ii}) \Lambda^T$ be the block modified Cholesky decomposition of Θ , and transform L' into $M' \in L'^+, \frac{m}{d}$ by $M' = L' \Lambda^{-T}$. The density of M' is then given by

$$(B.9) \quad \frac{1}{2^{\frac{1}{2}(mn+\frac{1}{2}(m^2-3m-2\tilde{m}))} \Gamma_{\frac{m}{d}}(\frac{1}{2}n))} \prod_{i=1}^d |\Delta_{ii}|^{-\frac{1}{2}n} |M'|^n \exp(-\frac{1}{2} \sum_{i \leq j} \text{tr}(M'_{ij} \Delta_{jj}^{-1} M'^T_{ij}))$$

with respect to $\nu_r(dM')$, that is $M' \in W_{\frac{m}{d}}^{\frac{1}{2}}(n, \bigoplus_{i=1}^d \Delta_{ii})$. Since $\Lambda \in L_{\frac{m}{d}}^I$ we have that $L'_{ii} = M'_{ii}$. By using normal theory when integrating out M'_{ij} for $i < j$, the joint distribution of M'_{ii} , $i=1, \dots, d$, is obtained as

$$\begin{aligned} (2^{\frac{1}{2}(mn-\frac{1}{2}(3m+m^2))} \pi^{\frac{1}{2}(\tilde{m}-m^2)}) \Gamma_{\frac{m}{d}}(\frac{1}{2}n))^{-1} & \left(\prod_{i=1}^d |\Delta_{ii}|^{-\frac{1}{2}(n-m_{i-1}^+)} |M'_{ii}|^n \right) \\ & \exp(-\frac{1}{2} \sum_{i=1}^d \text{tr}(M'_{ii} \Delta_{ii}^{-1} M'^T_{ii})) \end{aligned}$$

with respect to $\prod_{i=1}^d (dM'_{ii} / |M'_{ii}|^{m_i^+})$. Rewriting this density as

$$\begin{aligned} (B.10) \quad & \prod_{i=1}^d (2^{\frac{1}{2}(m_i(n-m_{i-1}^+)+\frac{1}{2}(m_i^2-3m_i-2m_i^2))} \Gamma_{m_i}(\frac{1}{2}(n-m_{i-1}^+)))^{-1} \\ & \left(\prod_{i=1}^d |\Delta_{ii}|^{-\frac{1}{2}(n-m_{i-1}^+)} |M'_{ii}|^{n-m_{i-1}^+} \right) \exp(-\frac{1}{2} \sum_{i=1}^d \text{tr}(M'_{ii} \Delta_{ii}^{-1} M'^T_{ii})) \end{aligned}$$

with respect to $\prod_{i=1}^d (dM'_{ii} / |M'_{ii}|^{m_i})$ we see that the M'_{ii} 's are independently distributed with $M'_{ii} \in W_{m_i}^{1/2}(n-m_{i-1}^+, \Delta_{ii})$. This proves (iii).

(iv) The result follows from (iii) and Lemma B.3. □

Parts of the proof follow the lines of the proof given by Eaton (1983) for the case $\underline{m}_d = \underline{1}_m$.

APPENDIX C: Proofs of Theorems 4.3, 4.5 and 4.6

PROOF OF THEOREM 4.3. The Fisher consistency follows from the fact that the estimates of $\phi(k)$ and $\Sigma(k)$ as functions of $G_{ij}(k)$, $i, j=1, \dots, p_k$, take the values $\phi(k)$ and $\Sigma(k)$ respectively at $\Gamma_{ij}(k)$, $i, j=1, \dots, p_k$.

Let $E_k = (e_k : e_{k+d} : \dots : e_{k+(N_k-1)d})$ where $e_k = y_k - \phi(k)y'_k$. We have $Y_k = \phi(k)Y'_k + E_k$ and $G_0(k) = \hat{\phi}^{YW}(k)G_{**}(k)$ and hence

$$(C.1) \quad \sqrt{N_k} (\hat{\phi}^{YW}(k) - \phi(k))G_{**}(k) = E_k Y'_k{}^T / \sqrt{N_k}$$

The sum of expected values of the squares of the elements of $E_k Y'_k{}^T$ is

$$\begin{aligned} \text{tr}(E(E_k Y'_k{}^T Y_k E_k{}^T)) &= E\left(\sum_{i=1}^{N_k} \sum_{j=1}^{N_k} e_{k+(i-1)d}^T e_{k+(j-1)d} y_{k+(i-1)d} y_{k+(j-1)d}'\right) \\ &= E\left(\sum_{i=1}^{N_k} e_{k+(i-1)d}^T e_{k+(i-1)d} y_{k+(i-1)d} y_{k+(i-1)d}'\right) \\ &= N_k \text{tr}(\Sigma(k)) \text{tr}(\Gamma_{**}(k)). \end{aligned}$$

Chebyshev's inequality now gives that

$$(C.2) \quad E_k Y'_k{}^T / N_k = (Y_k Y'_k{}^T - \phi(k) Y'_k Y_k{}^T) / N_k \xrightarrow{P} 0$$

that is $G_0(k) - \phi(k)G_{**}(k) \xrightarrow{P} 0$.

We have

$$\begin{aligned} (C.3) \quad \text{plim}_{N_k \rightarrow \infty} \hat{\phi}^{YW}(k) - \phi(k) &= \text{plim}_{N_k \rightarrow \infty} (Y_k Y'_k{}^T / N_k - \phi(k) Y'_k Y_k{}^T / N_k) (Y'_k Y_k{}^T / N_k)^{-1} \\ &= \text{plim}_{N_k \rightarrow \infty} (E_k Y'_k{}^T / N_k) (Y'_k Y_k{}^T / N_k)^{-1} = 0 \end{aligned}$$

by (C.2), that is $\hat{\phi}(k) \xrightarrow{P} \phi(k)$.

From the law of large numbers it follows that $E_k E_k{}^T / N_k \xrightarrow{P} 0$ as $N_k \rightarrow \infty$, that is

$$(C.4) \quad (Y_k Y'_k{}^T - \phi(k) Y'_k Y_k{}^T - Y_k Y'_k{}^T \phi(k)^T + \phi(k) Y'_k Y_k{}^T \phi(k)^T) / N_k \xrightarrow{P} \Sigma(k)$$

as N_k tends to infinity. By combining (C.4) with (C.2), suitably multiplied by $\phi(k)$, we infer that

$$(Y_k Y'_k{}^T - \phi(k) Y'_k Y_k{}^T \phi(k)^T) / N_k \xrightarrow{P} \Sigma(k)$$

that is $G_{00}(k) - \Phi(k)G_{11}(k)\Phi(k)^T \xrightarrow{P} \Sigma(k)$ as $N_k \rightarrow \infty$. Then $\hat{\Sigma}^{YW}(k) \xrightarrow{P} \Sigma(k)$ follows from

$$\begin{aligned} \hat{\Sigma}^{YW}(k) &= G_{00}(k) - G_{01}(k)G_{11}(k)^{-1}G_{10}(k) \\ (C.5) \quad &= (Y_k - \hat{\Phi}^{YW}(k)Y'_k)(Y_k - \hat{\Phi}^{YW}(k)Y'_k)^T/N_k \end{aligned}$$

and (C.3) and (C.4).

We have

$$(C.6) \quad E(E_k Y'_k Y_k^T) = E\left(\sum_{i=1}^{N_k} e_{k+(i-1)d} y_{k+(i-1)d}^T\right) = 0$$

and

$$\begin{aligned} E(E_k Y'_k Y_k^T \otimes E_k Y'_k Y_k^T) &= E\left(\sum_{i=1}^{N_k} \sum_{j=1}^{N_k} e_{k+(i-1)d} e_{k+(j-1)d}^T \otimes y_{k+(i-1)d} y_{k+(j-1)d}^T\right) \\ &= \sum_{i=1}^{N_k} E(e_{k+(i-1)d} e_{k+(i-1)d}^T) \otimes E(y_{k+(i-1)d} y_{k+(i-1)d}^T) \\ (C.7) \quad &= N_k \Sigma(k) \otimes \Gamma_{11}(k). \end{aligned}$$

To show the limiting distribution of $\sqrt{N_k}(\hat{\Phi}(k) - \Phi(k))$, it suffices by (C.1), (C.6) and (C.7) to show that $G_{11}(k) \rightarrow \Gamma_{11}(k)$ as $N_k \rightarrow \infty$ and that

$$E_k Y'_k Y_k^T / \sqrt{N_k} \xrightarrow{L} N_{m,mp_k}(0, \Sigma(k), \Gamma_{11}(k))$$

as $N_k \rightarrow \infty$.

We have that

$$G_{00}(j) = N_j^{-1} \sum_{i=1}^{N_j} y_{j+(i-1)d} y_{j+(i-1)d}^T$$

is the j th diagonal block of $N^{-1} \sum_{k=1}^N x_k x_k^T$ (or $(N+1)^{-1} \sum_{k=1}^{N+1} x_k x_k^T$) and likewise $G_{ab}(j)$, $a, b = 1, \dots, p_j$, are appropriate blocks of $N^{-1} \sum_{k=1}^N x_k x_k^T$, (if $p_j \leq d$) or sums of blocks of $(N/n')^{-1} \sum_{k=1}^{N/n'} z_k z_k^T$ (where z_k is given by (3.11)) if $p_j > d$. The latter sum converges in probability to the expected value of $z_k z_k^T$, that is W_n , and the blocks converge to the corresponding blocks of W_n , which are the expected values of $G_{ab}(j)$ for different a and b and j .

The asymptotic normality of $E_k Y'_k Y_k^T$ is shown in a similar manner as in

the stationary case, cf. Hannan (1970, p. 330). Crucial for the validity of the proof is the existence of an infinite moving average representation

$$y_k = \sum_{j=0}^{\infty} \phi_j(k) e_{k-j}.$$

The result follows from Theorem 3.1 and the expansion for the deperiodized process. Here $\phi_j(k) = \phi_j(k+d)$. □

PROOF OF THEOREM 4.5. The Yule-Walker estimates $\hat{\Psi}^{YW}$ and $\hat{\Theta}^{YW}$ are the solutions of the equations $C_{10} = C_{00}(\hat{\Psi}^{YW})^T$ and $C_{00} = C_0(\hat{\Psi}^{YW})^T + \hat{\Theta}^{YW}$. Let $Q = (X'X'^T)^{-}X'$ for some g-inverse $(X'X'^T)^{-}$ of $X'X'^T$, then $(\hat{\Psi}^{YW})^T = QX^T$ and $N\hat{\Theta}^{YW} = XPX^T$ where $P = (I_N - X'^TQ) = (I_N - Q^TX') = (I_N - X'^T(X'X'^T)^{-}X')$.

Since P is idempotent we have $N\hat{\Theta}^{YW} = XPPX^T$. A decomposition of $XPPX^T$ into LDL^T according to Lemma 3.1 corresponds to a decomposition of PX^T into $UD^{\frac{1}{2}}L^T$ according to Theorem B.2 in Appendix B. By using the block Gram-Schmidt orthogonalization procedure, we see that the elements $\hat{\Sigma}^{dYW}(k)$, $k=1, \dots, d$, in the decomposition $\hat{\Lambda}(\bigoplus_{k=1}^d N\hat{\Sigma}^{dYW}(k))\hat{\Lambda}^T$ of $N\hat{\Theta}^{YW}$ are given by

$$(C.8) \quad \hat{\Sigma}^{dYW}(k) = \hat{E}_k^T \hat{E}_k / N = Y_k (P - \sum_{i=1}^{k-1} \hat{E}_i (\hat{E}_i^T \hat{E}_i)^{-} \hat{E}_i^T) Y_k^T / N$$

for $k=1, \dots, d$. Here we have also used the fact that $P\hat{E}_i = \hat{E}_i$, which is easily verified by induction. It also follows from Theorem B.2 that

$$\hat{\Lambda}_{ij} = Y_i \hat{E}_j (\hat{E}_j^T \hat{E}_j)^{-} = (Y_i (P - \sum_{k=1}^{j-1} \hat{E}_k (\hat{E}_k^T \hat{E}_k)^{-} \hat{E}_k^T) Y_j^T) (\hat{E}_j^T \hat{E}_j)^{-}.$$

By inverting $\hat{\Lambda}$ into $\hat{\Lambda}^{-1} = (\hat{\Lambda}^{ij})$, say, we have that

$$\hat{\Lambda}^{ij} = -(\sum_{k=j+1}^i \hat{\Lambda}^{ik} Y_k \hat{E}_j) (\hat{E}_j^T \hat{E}_j)^{-},$$

recursively defined for $j=i-1, i-2, \dots, 1$ and $i=2, \dots, d$, and further $L\hat{\Lambda}^{ii} = I_{m_i}$ and $\hat{\Lambda}^{ij} = 0$ for $i < j$. Inserting $\hat{\Phi}_{i-j}^{dYW}(i)$ for $\hat{\Lambda}^{ij}$ and rewriting this expression, we obtain the expression for $\hat{\Phi}_j^{dYW}(k)$ for $j=1, \dots, k-1$.

We have $\hat{\Psi}^{YW} = XQ^T$ and by defining $(\hat{\Psi}^{YW})^T = (\hat{\Psi}_1^T, \dots, \hat{\Psi}_d^T)$ we obtain $\hat{\Psi}_k^T = Y_k Q^T$ for $k=1, \dots, d$. With $\hat{F}^T = (\hat{F}_1^T, \dots, \hat{F}_d^T)$, we have from $\hat{F} = \hat{\Lambda}^{-1} \hat{\Psi}^{YW}$ that

$$\begin{aligned}\hat{F}_{k\cdot} &= \hat{\Psi}_{k\cdot} - \sum_{i=1}^{k-1} L^{ki} \hat{\Psi}_{i\cdot} = Y_k Q^T - \sum_{i=1}^{k-1} \hat{\Phi}_{k-i}^{dYW}(k) Y_i Q^T \\ &= (Y_k - \sum_{i=1}^{k-1} \hat{\Phi}_i^{dYW}(k) Y_{k-i}) Q^T\end{aligned}$$

The blocks of $\hat{F}_{k\cdot}$ are $\hat{\Phi}_i^{dYW}(k)$, for $i = d+k-1, d+k-2, \dots, k, 2d+k-1, 2d+k, \dots, d+k, \dots, pd+k-1, \dots, (p-1)d+k$. By rearranging the order of the blocks by postmultiplying with $I_{md} \otimes R_p$, the order changes to $i = pd+k-1, pd+k-2, \dots, k$. Thus by postmultiplication with $I_m \otimes 1_{pd, pd+k-j}$ we obtain $\hat{\Phi}_j^{dYW}(k)$. This shows the expression for $\hat{\Phi}_j^{dYW}(k)$ for $j = k, k+1, \dots, pd+k-1$ and completes the proof of the theorem. $\square$

PROOF OF THEOREM 4.6. The motivation for the Fisher consistency is the same as in the proof of Theorem 4.3. An application of a result for stationary MAR processes (cf. Hannan 1970, Theorem 1, p. 329) shows that, for the process given by (3.1), $C_{n\cdot}$, $\hat{\Psi}$ and $\hat{\Theta}$ converge almost surely to $T_{n\cdot}$, Ψ and Θ , respectively. Moreover, the same theorem yields that

$$\sqrt{N} (\hat{\Psi} - \Psi) \xrightarrow{L} N_{md, mdp}(0, \Theta, T_{n\cdot}^{-1}) \text{ as } N \rightarrow \infty.$$

The rest of the theorem consists of the following steps:

The asymptotic normality of $\hat{\Theta}$ follows from the representation $\hat{\Theta} = (X - \hat{\Psi}X')(X - \hat{\Psi}X')^T/N$ by using that $\hat{\Psi}$ is consistent for Ψ . The limiting distributions of $\sqrt{N}\hat{\Theta}$ and $(X - \Psi X')(X - \Psi X')^T/\sqrt{N}$ are the same. Since it is true that

$$(X - \Psi X')(X - \Psi X')^T/\sqrt{N} \in W_{md}(N, \Theta/\sqrt{N})$$

and hence

$$\text{vec}((X - \Psi X')(X - \Psi X')^T/\sqrt{N} - \sqrt{N}\hat{\Theta}) \xrightarrow{L} N_{m^2 d^2}(0, 2S_{md} \Theta^{<2>} S_{md})$$

we obtain the result

$$\sqrt{N} \text{vec}(\hat{\Theta} - \Theta) \xrightarrow{L} N_{m^2 d^2}(0, 2S_{md} \Theta^{<2>} S_{md})$$

where $S_m = \frac{1}{2}(I_{m,m} + T_{m,m})$.

The approximation

$$(C.9) \quad N\hat{\Theta} \in W_{md}(N, \Theta)$$

is valid for large N . From Theorem B.1 (i) and (xi) in Appendix B, it then follows that the approximations

$$N \hat{\Sigma}^{dYW}(j) \in W_m(N-m(j-1), \Sigma(j)),$$

for $j=1, \dots, d$, and

$$(C.10) \quad \sqrt{N} \hat{\Lambda}_j^T \in T_{m, m(d-j)}(N-m(j-1), \sqrt{N} \Lambda_j^T, \Sigma(j)^{-1}, \Lambda_j'' (\bigoplus_{k=j+1}^d \Sigma(k)) \Lambda_j''^T)$$

for $j=1, \dots, d-1$, are valid for large N . Hence we see that

$$\sqrt{N} \text{vec}(\hat{\Sigma}^{dYW}(j) - \Sigma(j)) \xrightarrow{L} N_m^2(0, 2S_m \Sigma(j)^{<2>} S_m),$$

for $j=1, \dots, d$, and

$$(C.11) \quad \sqrt{N}(\hat{\Lambda}_j^T - \Lambda_j^T) \xrightarrow{L} N_{m, m(d-j)}(0, \Sigma(j)^{-1}, \Lambda_j'' (\bigoplus_{k=j+1}^d \Sigma(j)) \Lambda_j''^T),$$

for $j=1, \dots, d-1$, as $N \rightarrow \infty$.

From (C.11) it follows that $\hat{\Lambda}_j \xrightarrow{P} \Lambda_j$, $j=1, \dots, d-1$ and thus $\hat{\Lambda} \xrightarrow{P} \Lambda$ as $N \rightarrow \infty$. The asymptotic distribution of $\hat{F} = \hat{\Lambda}^{-1} \hat{\Psi}$ is the same as the asymptotic distribution of $\Lambda^{-1} \hat{\Psi}$. Hence the result (vi) follows from (i). Finally, we have the general relations $\hat{\Lambda}_j'' (\hat{\Lambda}^{-1})_{,j} = -\hat{\Lambda}_{,j}$ for $j=1, \dots, d-1$, between the blocks of $\hat{\Lambda}$ and the blocks of the inverse $\hat{\Lambda}^{-1}$. Since, according to (iii), it is true that $\hat{\Lambda}_j'' \xrightarrow{P} \Lambda_j''$, the asymptotic distribution of $(\hat{\Lambda}^{-1})_{,j}$ is the same as the asymptotic distribution of $-\Lambda_j''^{-1} \hat{\Lambda}_{,j}$; as follows from (iii), the latter is

$$N_{m, m(d-j)}(0, \bigoplus_{k=j+1}^d \Sigma(j), \Sigma(j)^{-1}).$$

This completes the proof of the theorem. □

APPENDIX D: Proof of Theorem 4.7

LEMMA D.1. Let X be a nonsingular m -square symmetric matrix, Y an arbitrary matrix and $S_m = \frac{1}{2}(I_{m,m} + T_{m,m})$. Then the following results hold:

- (i) $d_X \square A^T X^{-1} A = -(A^T X^{-1} \otimes A^T X^{-1}) S_m$
- (ii) $d_X \square \text{tr}(A^T X^{-1} A) = (\text{vec}(X^{-1} A A^T X^{-1}))^T S_m$
- (iii) $d_X \square \log |X| = (\text{vec}(X^{-1}))^T S_m$
- (iv) $d_Y \square \text{tr}(A^T Y B) = (\text{vec}(A B^T))^T$
- (v) $d_Y \square \text{tr}(A^T Y^T B) = (\text{vec}(B A^T))^T$
- (vi) $d_Y \square \text{tr}(A^T Y^T X Y B) = (\text{vec}(X Y B A^T + X Y A B^T))^T$
- (vii) $d_Y \square A Y B = A \otimes B^T$
- (viii) $d_Y \square A Y^T B = A \otimes B^T$

PROOF. The results (iv), (v), (vii) and (viii) are proved by Holmquist (1985b, Lemma 5.1). The result (i) follows from Theorem 5.3 (iii) by Holmquist (1985b) and from the correspondence between gradients with respect to matrices with linear dependence between the elements and the corresponding gradients without these restrictions on the elements, cf. Holmquist (1985b, Sec. 5.4). We have

$$d_X \square A^T X^{-1} A = -(A^T \otimes A^T)(X^{-1} \otimes X^{-1}) S_m = -(A^T X^{-1} \otimes A^T X^{-1}) S_m$$

The results (ii) and (iii) are shown similarly. The result (vi) follows from (iv) and (v) by using the decomposition rule.

PROOF OF THEOREM 4.7. The conditional log likelihood is

$$\begin{aligned} \ln f &= \ln f_{(1,n) | (-q,0)} = -\frac{1}{2} m n \ln(2\pi) \\ &\quad - \frac{1}{2} \sum_{k=1}^d \left(N_k \ln(|\Sigma(k)|) + \text{tr}((Y_k - \Phi(k) Y'_k)^T \Sigma(k)^{-1} (Y_k - \Phi(k) Y'_k)) \right). \end{aligned}$$

From Lemma D.1 (i) and (ii) we see that

$$\begin{aligned} \text{(D.1)} \quad d_{\Sigma(k)} \square \ln f &= -\frac{1}{2} (\text{vec}(N_k \Sigma(k)^{-1} \\ &\quad - \Sigma(k)^{-1} (Y_k - \Phi(k) Y'_k) (Y_k - \Phi(k) Y'_k)^T \Sigma(k)^{-1}))^T S_m \end{aligned}$$

and from Lemma D.1 (vi) and (vii) we infer that

$$(D.2) \quad d_{\Sigma(k)} \square (d_{\Sigma(k)} \square \ln f) = -\frac{1}{2} S_m (-N_k \Sigma(k)^{-1} \otimes \Sigma(k)^{-1} \\ + \Sigma(k)^{-1} (Y_k - \Phi(k) Y'_k) (Y_k - \Phi(k) Y'_k)^T \Sigma(k)^{-1} \otimes \Sigma(k)^{-1} \\ + \Sigma(k)^{-1} \otimes \Sigma(k)^{-1} (Y_k - \Phi(k) Y'_k) (Y_k - \Phi(k) Y'_k)^T \Sigma(k)^{-1}) S_m$$

From (D.1) we see that

$$(D.3) \quad d_{\Sigma(j)} \square (d_{\Sigma(k)} \square \ln f) = 0$$

for $j \neq k$ and

$$(D.4) \quad d_{\Phi(j)} \square (d_{\Sigma(k)} \square \ln f) = 0,$$

while

$$(D.5) \quad d_{\Phi(k)} \square (d_{\Sigma(k)} \square \ln f) = -S_m (\Sigma(k)^{-1} \otimes \Sigma(k)^{-1} (Y_k - \Phi(k) Y'_k) Y'_k{}^T).$$

Further,

$$(D.6) \quad d_{\Phi(k)} \square \ln f = (\text{vec}(\Sigma(k)^{-1} (Y_k - \Phi(k) Y'_k) Y'_k{}^T))^T$$

and thus for $j \neq k$,

$$(D.7) \quad d_{\Phi(j)} \square (d_{\Phi(k)} \square \ln f) = 0,$$

$$(D.8) \quad d_{\Phi(k)} \square (d_{\Phi(k)} \square \ln f) = -(\Sigma(k)^{-1} \otimes Y'_k Y'_k{}^T),$$

$$(D.9) \quad d_{\Sigma(j)} \square (d_{\Phi(k)} \square \ln f) = 0,$$

and

$$(D.10) \quad d_{\Sigma(k)} \square (d_{\Phi(k)} \square \ln f) = -(\Sigma(k)^{-1} \otimes ((Y_k - \Phi(k) Y'_k) Y'_k{}^T)^T \Sigma(k)^{-1}) S_m$$

Taking expectations, and using the facts that $E((Y_k - \Phi(k) Y'_k) (Y'_k - \Phi(k) Y'_k)^T) = N_k \Sigma(k)$, $E(Y'_k Y'_k{}^T) = N_k \Gamma_{\pi}(k)$ and $E((Y_k - \Phi(k) Y'_k) Y'_k{}^T) = 0$, we obtain from (D.2), (D.5), (D.10) and (D.8),

$$I(\Sigma(k), \Sigma(k)) = \frac{1}{2} N_k S_m (\Sigma(k)^{-1} \otimes \Sigma(k)^{-1}) S_m$$

$$I(\Sigma(k), \Phi(k)) = (I(\Phi(k), \Sigma(k)))^T = 0$$

$$I(\Phi(k), \Phi(k)) = N_k \Sigma(k)^{-1} \otimes \Gamma_{\pi}(k).$$

From (D.3), (D.4), (D.7) and (D.9) we have that $I(\Sigma(k), \Sigma(j))$,

$I(\Sigma(k), \Phi(j))$, $I(\Phi(k), \Sigma(j))$ and $I(\Phi(k), \Phi(j))$ are all zeromatrices of proper order if $j \neq k$. □

APPENDIX E: Proofs of Theorems 6.1 and 6.2

LEMMA E.1. Let $\{w_i\}_{i=-\infty}^{\infty}$ be a sequence of independent and identically distributed vector variables with $E(w_i) = 0$ and $E(w_i w_i^T) = \Sigma$. Define for fixed integers $d \neq 0$ and h ,

$$a_h = N^{-1} \sum_{i=1}^N w_{id} \otimes w_{id-h};$$

then if $h \neq 0$, $N^{\frac{1}{2}} a_h$ has a limiting normal distribution as $N \rightarrow \infty$ with zero mean and dispersion matrix $\Sigma \otimes \Sigma$. If, in addition, $E((w_i^T w_i)^2) < +\infty$, then $N^{\frac{1}{2}}(a_0 - \text{vec}(\Sigma))$ has a limiting normal distribution as $N \rightarrow \infty$ with zero mean and dispersion matrix $E(w_i^{<2>} w_i^{<2>T}) - \Sigma \square \Sigma$.

PROOF. Define $z_{i,h} = w_{id} \otimes w_{id-h}$. For $h \bmod d \neq 0$, the quantities $z_{1,h}, z_{2,h}, \dots, z_{N,h}$ are mutually independent and $E(z_{i,h}^T z_{i,h}) = (E(w_i^T w_i))^2 < +\infty$; thus a multivariate central limit theorem gives that $N^{\frac{1}{2}} a_h = N^{-\frac{1}{2}} \sum_{i=1}^N z_{i,h}$ has a limiting normal distribution with expectation

$$N^{-\frac{1}{2}} \sum_{i=1}^N E(z_{i,h}) = N^{-\frac{1}{2}} \sum_{i=1}^N E(w_{id}) \otimes E(w_{id-h}) = 0$$

and dispersion

$$\begin{aligned} N^{-1} \sum_{i=1}^N \sum_{j=1}^N E(z_{i,h} z_{j,h}^T) &= N^{-1} \sum_{i=1}^N E(w_{id} w_{id}^T \otimes w_{id-h} w_{id-h}^T) \\ &= E(w_{id} w_{id}^T) \otimes E(w_{id-h} w_{id-h}^T) = \Sigma \otimes \Sigma. \end{aligned}$$

For $h = 0$, a similar argument applies but the expectation and dispersion are in this case obtained by noting that

$$N^{-\frac{1}{2}} \sum_{i=1}^N E(z_{i,0}) = N^{-\frac{1}{2}} \sum_{i=1}^N E(w_{id}^{<2>}) = N^{\frac{1}{2}} \text{vec}(\Sigma)$$

and

$$\begin{aligned} N^{-1} \sum_{i=1}^N \sum_{j=1}^N E(z_{i,0} z_{j,0}^T) &= N^{-1} \sum_{i=1}^N \sum_{j=1}^N E(w_{id} w_{jd}^T \otimes w_{id} w_{jd}^T) \\ &= N^{-1} ((N^2 - N) E(w_{id}^{<2>}) \otimes E(w_{jd}^{<2>T}) + N E(w_{id}^{<2>} w_{id}^{<2>T})) \\ &= (N - 1) (\Sigma \square \Sigma) + E(w_{id}^{<2>} w_{id}^{<2>T}). \end{aligned}$$

For fixed u in $\{\pm 1, \pm 2, \dots\}$ and $h = ud$, the sequence $z_{1,h}, z_{2,h}, \dots$ is u -dependent. By application of a limit theorem for such processes, see

e.g. Diananda (1953, Theorem 2.A), we find that $N^{-\frac{1}{2}} \sum_{i=1}^N z_{i,h}$ has a limiting normal distribution with mean

$$N^{-\frac{1}{2}} \sum_{i=1}^N E(z_{i,ud}) = N^{-\frac{1}{2}} \sum_{i=1}^N E(w_{id}) \otimes E(w_{(i-u)d}) = 0$$

and dispersion

$$\begin{aligned} N^{-1} \sum_{i=1}^N \sum_{j=1}^N E(z_{i,ud} z_{j,ud}^T) &= E(z_{1,ud} z_{1,ud}^T) + E(z_{1,ud} z_{2,ud}^T) \\ &+ E(z_{2,ud} z_{1,ud}^T) + \dots + E(z_{1,ud} z_{u+1,ud}^T) + E(z_{u+1,ud} z_{1,ud}^T) \\ &= \Sigma \otimes \Sigma. \end{aligned}$$

The last equality follows from that $E(z_{i,ud} z_{j,ud}^T) = 0$ for $i \neq j$. $\square$

For $d = 1$, a_h is the usual autocovariance of lag h written in vector form.

LEMMA E.2. Let $\{v_i\}_{i=-\infty}^{+\infty}$ and $\{w_i\}_{i=-\infty}^{+\infty}$ be two mutually independent sequences each consisting of independent and identically distributed vector variables with $E(v_i) = E(w_i) = 0$, $E(v_i v_i^T) = \Sigma_v$ and $E(w_i w_i^T) = \Sigma_w$. If we define, for fixed integers $d \neq 0$ and h , the quantities

$$c_h = N^{-1} \sum_{i=1}^N v_{id} \otimes w_{id-h},$$

then $N^{\frac{1}{2}} c_h$ has a limiting normal distribution as $N \rightarrow \infty$ with zero mean and dispersion matrix $\Sigma_v \otimes \Sigma_w$.

PROOF. Define $z_{i,h} = v_{id} \otimes w_{id-h}$. Then the quantities $z_{1,h}, z_{2,h}, \dots, z_{N,h}$ are mutually independent and $E(z_{i,h}^T z_{i,h}) = E(v_i^T v_i) E(w_i^T w_i) < +\infty$. Thus by applying multivariate central limit theorem, we get that $N^{\frac{1}{2}} c_h = N^{-\frac{1}{2}} \sum_{i=1}^N z_{i,h}$ has a limiting normal distribution with expectation

$$N^{-\frac{1}{2}} \sum_{i=1}^N E(z_{i,h}) = N^{-\frac{1}{2}} \sum_{i=1}^N E(v_{id}) \otimes E(w_{id-h}) = 0$$

and dispersion

$$\begin{aligned} N^{-1} \sum_{i=1}^N \sum_{j=1}^N E(z_{i,h} z_{j,h}^T) &= N^{-1} \sum_{i=1}^N \sum_{j=1}^N E(v_{id} v_{jd}^T) \otimes E(w_{id-h} w_{jd-h}^T) \\ &= \Sigma_v \otimes \Sigma_w. \end{aligned} \quad \square$$

For $d = 1$, c_h is the cross-covariance function of lag h , written in vector form.

We are now able to prove Theorem 6.1.

PROOF OF THEOREM 6.1. (i) and (ii): The moments are obtained by algebraic manipulations and by using the independence of the sequence $\{e_k\}$. We derive the expression for $\mathcal{D}(S_{i,j}(k))$ for $i \neq j$, in which case we have

$$\begin{aligned} \mathcal{D}(S_{i,j}(k)) &= E(s_{i,j}(k) s_{i,j}(k)^T) \\ (E.1) \quad &= N_k^{-2} \sum_{u=1}^{N_k} \sum_{v=1}^{N_k} E(e_{k-i+(u-1)d} e_{k-i+(v-1)d}^T \otimes e_{k-j+(u-1)d} e_{k-j+(v-1)d}^T). \end{aligned}$$

Since for $i \neq j$, there exist no u, v such that $k-j+(u-1)d = k-j+(v-1)d$ and $k-i+(v-1)d = k-j+(u-1)d$ other than $u=v$. Hence (E.1) can be written

$$\begin{aligned} &N_k^{-2} \sum_{u=1}^{N_k} E(e_{k-i+(u-1)d} e_{k-i+(u-1)d}^T \otimes e_{k-j+(u-1)d} e_{k-j+(u-1)d}^T) \\ &N_k^{-2} \sum_{u=1}^{N_k} \Sigma(k-i) \otimes \Sigma(k-j) = N_k^{-1} \Sigma(k-i) \otimes \Sigma(k-j). \end{aligned}$$

The other expressions are shown similarly.

(iii): We have

$$\begin{aligned} &E(s_{i,j}(k) s_{i,l}(k)^T) \\ &= N_k^{-2} \sum_{u=1}^{N_k} \sum_{v=1}^{N_k} E(e_{k-i+(u-1)d} e_{k-i+(v-1)d}^T \otimes e_{k-j+(u-1)d} e_{k-l+(v-1)d}^T) \\ &= 0, \end{aligned}$$

since $k-i+(v-1)d \neq k-l+(v-1)d$ and there exist no pairs u, v such that $k-i+(u-1)d = k-l+(v-1)d$ and $k-i+(v-1)d = k-j+(u-1)d$.

(iv) and (v): For $(i-j) \bmod d \neq 0$ we have by a direct application of Lemma E.2 with $v_{ld} = e_{k-i+(\ell-1)d}$ and $w_{ld-(j-i)} = e_{k-j+(\ell-1)d}$ that $N_k^{\frac{1}{2}} s_{i,j}(k)$ has a limiting normal distribution with zero mean and dispersion matrix $\Sigma(k-i) \times \Sigma(k-j)$. For $j = i + ud$, for fixed u in $\{\pm 1, \pm 2, \dots\}$ we apply Lemma E.1 with $w_{ld} = e_{k-i+(\ell-1)d}$ and $w_{ld-ud} = e_{k-j+(\ell-1)d}$, which yields the result.

Finally for $j = i$, we again apply Lemma E.1 with $w_{\ell d} = e_{k-i+(\ell-1)d}$, which yields the result for this case. This completes the proof of the theorem. $\square$

PROOF OF THEOREM 6.2. The proof is similar to that of Theorem 6.1, although the dimension is higher. The moments of the processes $\{v_k\}$ and $\{u_k\}$ are easily expressed in terms of the moments of the process $\{e_k\}$. $\square$

APPENDIX F: Proofs of Theorems 6.3 and 6.4

PROOF OF THEOREM 6.3. It follows from Lemma 3 of Hosking (1980) that to $O_p(N^{-1})$, it hold for $h > p$ that

$$(F.1) \quad \text{vec}(\hat{U}_{12} : \dots : \hat{U}_{1,h+1}) = (I_{(md)^2 h} - Q) \text{vec}(U_{12} : \dots : U_{1,h+1})$$

where Q is a $(md)^2 h$ -square matrix of rank $(md)^2 p$. The matrix Q can be written $Q = Z(Z^T(\Theta \otimes \Theta \otimes I_h)^{-1}Z)^{-1}Z^T(\Theta \otimes \Theta \otimes I_h)^{-1}$ where

$$Z = \begin{pmatrix} G_0 & 0 & 0 & \dots & 0 \\ G_1 & G_0 & 0 & \dots & 0 \\ G_2 & G_1 & G_0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ G_{h-1} & G_{h-2} & G_{h-3} & \dots & G_{h-p} \end{pmatrix}$$

and $G_k = I_{md} \otimes \psi_k$ where the md -square matrix ψ_k is defined by the relation

$$\sum_{k=0}^{\infty} \psi_k z^k = -(\sum_{k=0}^p \psi_k z^k)^{-1};$$

where $\psi_0 = -I_m$, cf. Hosking (1980). By writing $Z = (I_{md} \otimes \Theta \otimes I_h)(I_{md} \otimes Z_o)$ where

$$Z_o = \begin{pmatrix} \psi_0 & 0 & 0 & \dots & 0 \\ \psi_1 & \psi_0 & 0 & \dots & 0 \\ \psi_2 & \psi_1 & \psi_0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \psi_{h-1} & \psi_{h-2} & \psi_{h-3} & \dots & \psi_{h-p} \end{pmatrix}$$

we have that $Q = I_{md} \otimes (\Theta \otimes I_h) Z_o (Z_o^T (\Theta \otimes I_h) Z_o)^{-1} Z_o^T = I_{md} \otimes Q_o$ where

$$Q_o = (\Theta \otimes I_h) Z_o (Z_o^T (\Theta \otimes I_h) Z_o)^{-1} Z_o^T$$

is a mdh -square idempotent matrix of rank mdp . Hence we may write (F.1) as

$$(F.2) \quad (\hat{U}_{12} : \dots : \hat{U}_{1,h+1}) = (U_{12} : \dots : U_{1,h+1}) (I_{mdh} - Q_o^T)$$

to $O_p(N^{-1})$. Since for large N , $(U_{12} : \dots : U_{1,h+1}) \in N_{md,mdh}(0, \Theta/N, \Theta \otimes I_h)$, (Theorem 6.2 (iii b) and (iv b)) we have from (F.2) that for large N

$$(\hat{U}_{12} : \dots : \hat{U}_{1,h+1}) \in N_{md,mdh}(0, \Theta/N, (\Theta \otimes I_h)(I_{mdh} - Q_o^T))$$

and

$$(\hat{U}_{12} \theta^{-1/2} : \dots : \hat{U}_{1,h+1} \theta^{-1/2}) \in N_{md,mdh}(0, \theta/N, I_{mdh} - Q_1^T)$$

where $Q_1 = (\theta^{1/2} \otimes I_h) Z_o (Z_o^T (\theta \otimes I_h) Z_o)^{-1} Z_o^T (\theta^{1/2} \otimes I_h)$ is an mdh-square orthogonal projection of rank mdp.

If a random matrix has the distribution $N_{m,n}(0, \Sigma, Q)$, where Q is an orthogonal projection of rank $r (\leq n)$, then $XX^T \in W_m(r, \Sigma)$ and $\Sigma^{1/2} XX^T \Sigma^{-1/2} \in W_m(r, I_m)$ (see Eaton (1983, Prop 8.5)). Hence

$$\begin{aligned} N(\hat{U}_{12} \theta^{-1/2} : \dots : \hat{U}_{1,h+1} \theta^{-1/2}) (\hat{U}_{12} \theta^{-1/2} : \dots : \hat{U}_{1,h+1} \theta^{-1/2})^T \\ = \sum_{k=2}^{h+1} N \hat{U}_{1k} \theta^{-1} \hat{U}_{1k}^T \in W_{md}(md(h-p), \theta) \end{aligned}$$

and

$$\begin{aligned} N \theta^{-1/2} (\hat{U}_{12} \theta^{-1/2} : \dots : \hat{U}_{1,h+1} \theta^{-1/2}) (\hat{U}_{12} \theta^{-1/2} : \dots : \hat{U}_{1,h+1} \theta^{-1/2})^T \theta^{-1/2} \\ = \sum_{k=2}^{h+1} N \theta^{-1/2} \hat{U}_{1k} \theta^{-1} \hat{U}_{1k}^T \theta^{-1/2} \in W_{md}(md(h-p), I_{md}) \end{aligned}$$

for large N .

The same result is obtained if θ is replaced by an consistent estimator such as $\hat{U}_{11}$, and hence the result follows. □

PROOF OF THEOREM 6.4. The proof follows the same lines as in the time homogeneous case, cf. Hosking (1980). Essential for the proof is the moving average representation

$$(F.3) \quad y_k = \sum_{\ell=0}^{\infty} \phi_{\ell}(k) e_{k-\ell}$$

where $\phi_{\ell}(k) = \phi_{\ell}(k+d)$. In $S_{ij}(k)$ are involved $y_{k-i}, y_{k-i+d}, \dots$ and $y_{k-j}, y_{k-j+d}, \dots$. For such series, the coefficient matrices in the expansion (F.3) do not vary within the series because of the periodicity of ϕ_{ℓ} . These coefficient matrices play the same role as the time independent coefficient matrices in the moving average expansion of time homogeneous autoregressive processes. □

APPENDIX G

LEMMA G.1. Let $X \in GL_m(\mathbb{R})$, $A \in \mathbb{R}^{m,m}$ and $c \in \mathbb{R} \setminus \{0\}$. Define the mapping $Q: GL_m(\mathbb{R}) \rightarrow \mathbb{R}$ by

$$Q(X) = c \log(|X|) + \text{tr}(AX^{-1}).$$

Then

(i) Q has stationary values if and only if $A \in GL_m(\mathbb{R})$

(ii) If $A \in GL_m(\mathbb{R})$, then there exists exactly one stationary mapping $\hat{X}$ given by

$$\hat{X} = A/c$$

(iii) If $A \in GL_m^+(\mathbb{R})$ ($GL_m^-(\mathbb{R})$) (the subspace of positive and negative definite matrices respectively), then Q has a minimum (maximum) at $\hat{X} = A/c$ if $c > 0$ and a maximum (minimum) if $c < 0$ and the extreme value is

$$Q(A/c) = c(\log(|A|) - \log(c) + m).$$

PROOF. Straightforward algebra gives (cf. Lemma D.1 in Appendix D or Holmquist (1985b, Lemma 5.1))

$$(G.1) \quad d_X \square (c \log(|X|) + \text{tr}(AX^{-1})) = (\text{vec}(cX^{-T} - X^{-T}A^T X^{-T}))^T.$$

If A is nonsingular, then $X = A/c$ (which is nonsingular) is a stationary matrix and also the only one. This shows (ii) and the sufficiency part of (i). On the other hand, if Q has stationary values these satisfy

$$cX^{-T} = X^{-T}A^T X^{-T}$$

which is equivalent to that $cX = A$; this also implies that A is nonsingular since X is. Hence the necessity part of (i) is proved

The second order differential is

$$(G.2) \quad (X^{-T}A \otimes I + I \otimes X^{-1}A - c)(X^{-T} \otimes X^{-1})_{T_{m,m}}$$

which at the stationary point is equal to

$$c^3(A^{-T} \otimes A^{-1})_{T_{m,m}}.$$

Let $Y = xy^T$ be a decomposable element in $\mathcal{R}^{m,m}$ then

$$\begin{aligned} & (\text{vec}(Y))^T (c^3 (A^{-T} \otimes A^{-1})_{m,m}) \text{vec}(Y) \\ &= c^3 (x^T \otimes y^T) (A^{-T} \otimes A^{-1}) (y \otimes x) = c^3 (y^T A^{-1} x)^2 \end{aligned}$$

which is nonnegative if $c > 0$ and nonpositive if $c < 0$. The same is true on the linear closure of decomposable elements, that is $\mathcal{R}^{m,m}$.

